Searching for Meaning in Gestural Data.

Interpretive Feature Extraction and Signal Processing for Affective and Expressive Content

Teresa Marrin Nakra
MIT Media Laboratory
teresa@immersionmusic.org

Introduction

During the past twenty years there have been tremendous innovations in the development of interfaces and methods for performing live music with computers. However, most of these systems (with a few notable exceptions) have not been widely adopted by performing musicians. One explanation for this lack of use is that these technologies do not yet convey the most deeply meaningful aspects of human expression. That is, they do not capture and communicate the significant and emotional aspects of the gestures that are used to control them. Perhaps this is why so few musicians have used computer technology to replace, enhance, or transform the capabilities of their guitars, violins, and conducting batons.

However, there is a very strong case to be made for the need for new instruments: music literacy rates are declining, as is attendance and support for our "classical" music forms. As the mass production of the modern piano transformed American music history by introducing recreational music into the homes of middle-class families during the nineteenth century, it is possible that a new means for creating music might attract a large segment of the technologically savvy populace of the twenty-first century.

Not only could novel instruments captivate the imaginations of vast masses of amateurs, but they might also change and update our perceptions of *art music* by inspiring performing artists to develop new forms for the stage that preserve the role of *music for its own sake* in our culture. This collection of new instruments could leverage off of available technologies to do more than just play notes; they could generate and vary complex patterns, perform higher-level functions like conducting, and even generate graphics, lighting, and special effects. But in order for the performing arts community to consider these possibilities, it must have instruments that are at the very least as expressive as the traditional, mechanical ones have been.

Another reason to improve the capabilities of our new instruments is to retain and continue our tradition of live performing arts; in this age of easily-available recorded music it is rare to experience music which is unique to the moment in which it is heard. Live performances have inherent value because they reflect the states of the artists, the instruments, the audience, and the venue.

All of these factors interact in complex ways and sometimes produce surprising and exciting results. From perspective of a musician, live performances allow him or her to make real-time choices that affect the interpretive variation in the music. These choices give the performance a trajectory and a range of variation that define the expressiveness and power of the performance. Techniques for creating this variation involve subtle control over aspects such as timing, volume, timbre, accents, and articulation – sometimes implemented on many levels simultaneously. Musicians intentionally apply these techniques in the form of time- varying modulations on the structures in the music in order to express feelings and dramatic ideas[1] – some of which are pre-rehearsed and some of which change based on their own moods and whims.

In order for skilled performers to intentionally modulate these musical lines in real-time, they require musical instruments that are not only sensitive to subtle variations in input but can be also used to control multiple modulation streams in real-time; these instruments must also be repeatable and deterministic in their output.

What often happens with the replacement of traditional mechanical instruments with sensor-based interfaces is that the many dimensions in the input stream are reduced in the transduction process, and thereby effectively projected down to minimal axes of control. While the reduction of dimensionality in an input stream is often good for automatic recognition and other engineering tasks, it is not always ideal for music, where subtlety in the 'microstructural variation'[2] provides much of the interesting content for the performer and the listener. Another big problem with many sensor-based systems is the large disconnect between the way the gesture looks and the way the music sounds. This makes it difficult for an audience to understand or be able to respond to the performance. As a result of brittle, unnatural, overly constrained, or unsatisfying mappings between gesture and sound, these instruments can quickly become frustrating to performers; these problems must be solved before sensor-based instruments will be widely adopted.

This article presents a novel method that has been developed in order to improve upon the usefulness and attractiveness of our computer-based musical instruments. The primary task was to gather, analyze, and interpret gestural and physiological data from performing musicians in real rehearsal and performance conditions. Numerous background and design considerations will first be presented, followed by the results of an extensive visual analysis. Implications for these results and future work will be discussed at the end.

Background

The issue of expressivity in electronic musical performance is not new; many others have explored it through research projects and performances during the past sixty years. A few important examples of the use of sensor-based instruments onstage include Clara Rockmore's legendary performances on the Theremin[3], performance artist Laurie Anderson's work with the BodySynth[4], and Tod Machover's *Brain Opera*[5]. Other well-known gestural instruments include the BioMuse[6] and Radio Drum[7], which were developed at Stanford's Center for Computer Research in Music and Acoustics, and Laetitia Sonami's Lady's Glove[8] that was developed at STEIM in Amsterdam. Beginning in 1987 at the MIT Media Lab, Professor Tod Machover and his students began to bring similar ideas much closer to the classical performing arts traditions with his *Hyperinstruments*[9] project. About his research, Machover wrote:

> "Enhanced human expressivity is the most important goal of any technological research in the arts. To achieve this, it is necessary to augment the sophistication of the particular tools available to the artist. These tools must transcend the traditional limits of amplifying human gestuality, and become stimulants and facilitators to the creative process itself."[10]

1. Manfred Clynes articulated this idea in several articles, including "Generative Principles of Musical Thought," (*CCAI, Journal for the Integrated Study of Artificial Intelligence, Cognitive Science and Applied Epistemology*, 1986, volume 3, number 3, pp. 185-223), and "Microstructural Musical Linguistics: Composers' Pulses are Liked by the Best Musicians." (Cognition, 1995, volume 55, pp. 269-310). In the former, Clynes writes that "the principles which we have found to unify microstructure and structure thus account for a large part of the meaningfulness of music. When the appropriate microstructure is not present we tend to imagine it in our thought when listening to music...two thirds of the information of the music resides in the microstructure."(p. 6) In the latter, he defines unnotated microstructure as consisting of: "(1) time deviations of a note from the value given in the score; (2) amplitude of individual notes; (3) amplitude envelope of individual notes; (4) vibrato; (5) timbre changes within an individual note."(pp. 270-271) In the former he also includes "pulse" and "essentic forms" in his definition of microstructure.
2. Manfred Clynes' term, as elucidated in "Some guidelines for the synthesis and testing of Pulse Microstructure in relation to musical meaning", *Music Perception*, 7 (1990), 4, pp. 403-422.
3. Chadabe, Joel. *Electric Sound: the Past and Promise of Electronic Music*, pp. 9-10. NJ: Prentice-Hall, 1997.
4. David Morgenstern, "Performance Art and Trashcans." *MicroTimes*, June 28, 1993, p. 128. Jon Pareles, "Vaudeville, Complete with a Tornado". *New York Times*, Saturday, April 17, 1993.
5. http://brainop.media.mit.edu.
6. Lusted, Hugh and Benjamin Knapp, "Music Produced by Human Bioelectric Signals." See bibliography for full reference.

Among the more popular and enduring of the resultant family of hyperinstruments have been the Hyperviolin, the Hypercello, and the Sensor Chair[11], all of which were designed for expert and practiced performers.

Another hyperinstrument that was designed to be used by a practiced performer is a musical conducting system called the Digital Baton[12], which was intended to capture and map gestures into sound in the way that an orchestra might interpret the movements of a conductor.

The Baton is a ten-ounce molded polyurethane device that incorporates eleven sensory degrees of freedom: 2 degrees of position, 3 orthogonal degrees of acceleration, and 5 points of pressure. Professor Machover incorporated the instrument into his *Brain Opera* performance system, where it was used to trigger and shape multiple layers of sound in the live, interactive show[13]. Having been the primary designer of the Digital Baton, I was also its primary performer for nearly 180 live performances all over the world. Despite the high hopes I had for the Digital Baton, however, it ultimately failed to match my expectations of expressivity. Perhaps because I had been given the opportunity to perform with it so often, I became acutely aware of its shortcomings. Its biggest problems were:

- it was too big and too heavy to allow for graceful, comfortable gestures;
- its size and shape caused the wrist to take an unnatural, fixed grip;
- it was much more difficult than I originally imagined to map the positional information to anything useful other than fixed two-dimensional grids. Perhaps the expressive information in conducting is not contained in the positional information;
- accelerational data also turned out to be problematic, since the accelerometers' signal strength decreases nonlinearly as it rotates off-axis from gravity. Beats can be extracted from that information with processing, but are not always reliable. It turned out that force seems to be a better measure of intensity than acceleration;
- the event model was too simple to be useful in extracting significant events out of the data. Significant, expressive events often resembled insignificant, unexpressive ones.
- I only knew how to write linear mappings; I didn't have higher-order, nonlinear event models, and couldn't distinguish continuous from symbolic events;
- the disembodiment problem: it often wasn't obvious to the audience how the baton was controlling the sound;
- the over-constrained gesture problem: brittle recognition algorithms sometimes forced the performer to make exaggerated gestures in order to achieve a desired musical effect.

---

7. Developed by Bob Boie, Max Mathews, and Andy Schloss. See bibliography for full reference.
8. Developed with Bert Bongers. See the article by Bean entitled "A Soft Touch", *Electronic Musician*, June 1998, pp. 106-109.
9. Machover, Tod and Joseph Chung. "Hyperinstruments: musically intelligent interactive performance and creativity systems." *Proceedings of the International Computer Music Conference*, 1989.
10. Machover, Tod. "Hyperinstruments: a Progress Report, 1987-1991." MIT Media Laboratory, 1992.
11. The Sensor Chair uses a novel design of electric field sensors to track gestures in 2_ dimensions, invented by Professors Joseph Paradiso and Neil Gershenfeld of the MIT Media Lab and described in "Musical Applications of Electric Field Sensing." *Computer Music Journal* 21:2, pp. 69-89, Summer 1997.
12. Designed and built by a team at the MIT Media Lab including Teresa Marrin, Professor Joseph Paradiso, Professor Tod Machover, Maggie Orth, Chris Verplaetse, Pete Rice, and Patrick Pelletier.
13. More information on the Brain Opera can be found at: http://brainop.media.mit.edu. The Digital Baton was further developed by Professor Joseph Paradiso and Kai-Yuh Hsiao; information on their work can be found at: http://www.media.mit.edu/~joep/SpectrumWeb/captions/Baton.html.

My first response to these problems was to attempt to formulate a generalized theory of mappings[14], which resulted in a scheme for categorizing gestures along successive layers of complexity. This allowed for creating sophisticated, high-level action-descriptions from a sequence of minute atoms and primitives, in much the same way that languages are constructed out of phonemes. Ultimately, however, I realized that theorizing about mappings would not solve the fundamental problems of the Digital Baton. Instead, these issues could only be solved by means of an in-depth, quantitative, signal-based approach.

The Conductor's Jacket project

Formulated to respond to the issues raised by the Digital Baton, the Conductor's Jacket[15] project was begun in the spring of 1997. It was designed as a series of experiments to outfit professional conductors with a variety of physiological and motion sensors in order to gain more insight into precisely what gestures and signals performers create in real-life musical conditions. It was hoped that quantitative data from such situations would yield insight into both the internal experiences and external expressions of musicians and possibly reflect their feelings and musical interpretations. Some of our fundamental assumptions were that conductors would be a good subset of musicians to study because:

- they are considered to be among the most skillful, expert, and expressive of all musicians
- they have to amplify their gestures in order to be easily seen by many people
- their gestures are not constrained by any external instrument; they have free motion of their upper body
- their actions influence the higher-level functions of music, such as tempo, dynamics, phrasing, and articulation
- they purposefully (if not self-consciously) modulate the apparent viscosity of the air around them in order to communicate expressive effects

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

Each subject was given a unique, fitted shirt, into which were sewn eight physiological sensors: four muscle tension (electromyography, or EMG) sensors, one breathing monitor, one heart rate monitor, one skin conductance (galvanic skin response, or GSR) sensor, and one temperature sensor. One of the professional subjects also wore eight motion capture sensors. Design specifics of the wearable sensing and data collection systems are described in detail elsewhere.[16] Figure 1, below, demonstrates the placement of the different sensors in the Conductor's Jacket;[17] different subjects had slightly different arrangements, but all closely resembled this image in figure 1 below.

One motivation for this design was that I felt that the interface should have a relationship to (and a dependence on) the form of the human body; it should feel natural to wear and allow the physical frame to move. Regarding issues of sensor placement and interface design, our primary concern was to capture enough degrees of freedom to detect the *quality* and *complexity* of our subjects' motions. Finally, the

14. See Chapter 4 in *Toward an Understanding of Musical Gesture: Mapping Expressive Intention with the Digital Baton*. M.S. Thesis, MIT Media Laboratory, June 1996. Available in html form at: http://www.media.mit.edu/~marri.
15. This project was suggested by Professor Rosalind Picard of the MIT Media Laboratory and supported by the Affective Computing and Hyperinstruments Research Groups.
16. Marrin, Teresa and Rosalind Picard, "Analysis of Affective Musical Expression with the Conductor's Jacket." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, 1998.
17. Marrin, Teresa and Rosalind Picard. "The Conductor's Jacket: a Device for Recording Expressive Musical Gestures." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, 1998, p. 216

wearable interface was designed so as to not interfere with or impede movement by being light and unobtrusively sewn into a normal garment. My goal was to prevent the sensing environment from constraining the range and style of movement that the performers made.

In addition to design concerns of wearability, safety, and comfort for our subjects, I also had to address practical issues of data acquisition and sensing hardware in order for the data to be useful. First of all, measures had to be taken to remove all possible sources of noise; this included using shielded cable and a physical wire connection between the subjects and our data acquisition system. In addition, I included a grounding strap for the subject and conservative protections against electrical charges injuring the subject. Secondly, sampling rates for the data were a concern, since undersampled data would contain aliasing and would not be usable for signal processing tasks. Since events happen extremely quickly and precisely in conducting, the required time-scale is extremely small. Electromyography studies have shown that muscular contractions contain frequencies up to ~450 Hz[18], so they must be sampled at 900 Hz in order to completely capture the signal and satisfy the Nyquist criteria. Early versions of the system were only able to sample at 330 Hz, but later improvements increased the sampling rate to 2 kHz per channel.

18.S. DeLuca, "Surface Electromyography," http://www.delsys.com.

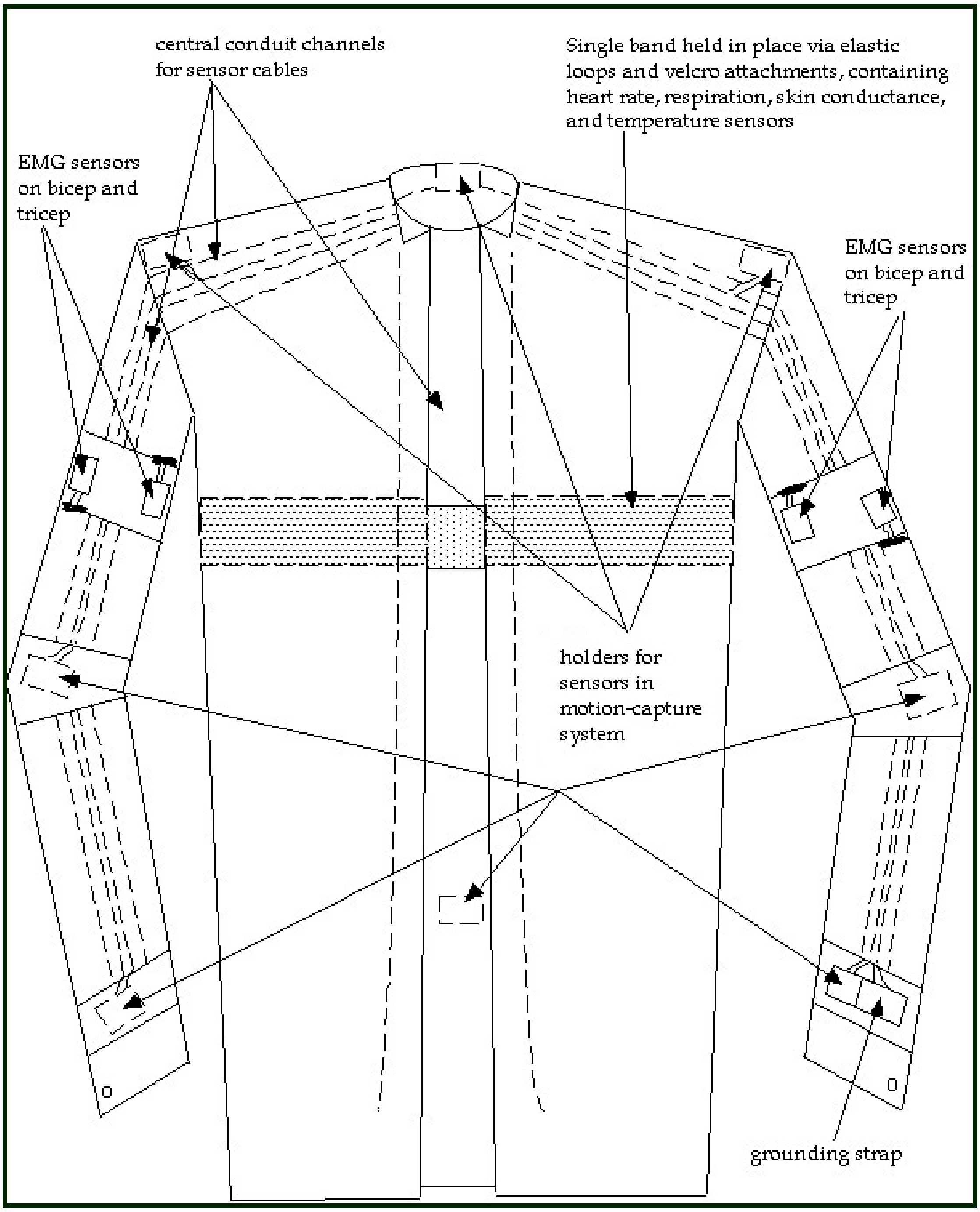

Fig. 1. Integration of physiological sensors into wearable form.

Interpretive Feature Identification

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

For the sake of this paper and other writings on this study, the subjects will be referred to simply by student/professional status and number (in order of when their data was collected). This is to minimize the effect of external associations that the reader might have and to avoid any unintended or unfortunate judgements or comparisons between subjects. Therefore, the six subjects will be mentioned simply as P1, P2, P3, and S1, S2, S3. Of the three professional subjects in this study, all three make a living from their musical performing and teaching in the Boston area. Two conduct mostly professional ensembles, while one conducts mostly amateur and student ensembles. All three permitted data recording during full rehearsals of their respective ensembles; one invited us to record data during a live performance. In all, 6 hours of data were collected from subject P1, 2 hours from subject P2, and 4 hours from subject P3.

The student subjects were all taking the Advanced Conducting course at Boston Conservatory; they had all had at least three years of conducting study. Their data was collected while they conducted a pianist and were simultaneously being evaluated by their instructor. All three conducted the same piece: Beethoven's Symphony no. 9, last movement. In all, S1 contributed 6 minutes' worth of data, S2 5 minutes, and S3 30 minutes. Most of the sensors provided reliable data throughout the experiments, although the heart-rate sensor failed to behave reliably and therefore its data has been discarded. Also, a full analysis of the positional data from subject P3 has not yet been completed, and therefore has not been included here.

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

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

Below is a thorough elaboration and illustration of these fourteen points; while they do not contain exhaustive proofs for each feature, the attempt has been to use cogent examples to elucidate the phenomenon, with enough detail that these results could be tested with more extensive research.

1. Use of the left hand for expressive variation

One of the first features that leapt out of the data was the very clear separation between conductors' use of their right and left arms. Since EMG is a measure of muscle *tension*, not necessarily movement, the EMG signal elucidates when the arm becomes engaged and actively generating signals. Therefore, it sometimes yields surprising results. Traditional conducting pedagogy teaches that the left hand should be used to provide supplementary information and expression; the EMG signals often supported this. P1's EMG signals demonstrate very clearly how he uses his left hand for extra emphasis; in the first two examples, below, EMG signals from P1's right and left biceps demonstrate how the left hand was used to supplement the information given by the right hand. In the first example, P1 modulated the meter from 2 beats per measure to 1 beat per measure. At the moment just before he intended to change the meter, he reached out his left hand (which was until that moment at his side) and reinforced the new meter with both hands. Figure 2, shown below, demonstrates how the previous faster meter (where only the right hand was used) transitioned to a slower meter as the left hand entered:[19]

19.Marrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 62.

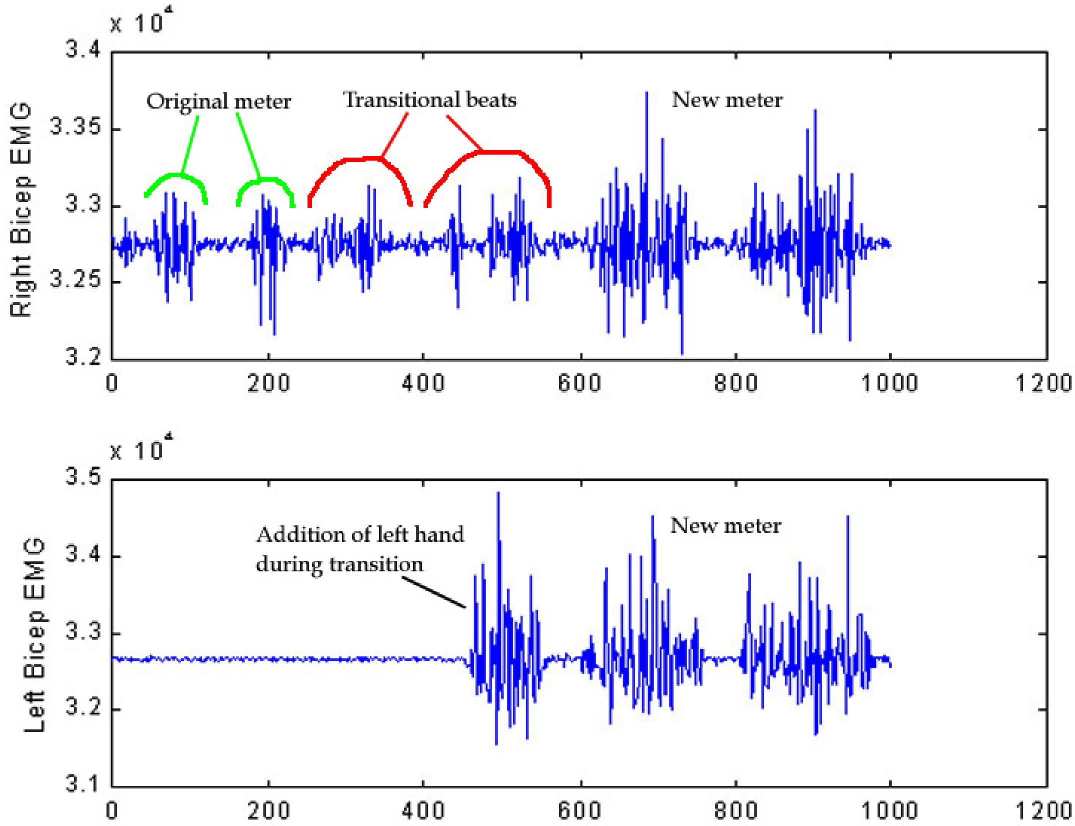

Fig. 2. EMG signals from both biceps during a metrical shift.

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

In another segment of the same session, subject P1 used his left hand to indicate a drastic reduction in loudness at the very end of a movement. As shown in figure 3, below, the right hand gave all the beats leading up to the ending, but at the last moment the left hand was raised (as the right hand was withdrawn) to indicate a quick volume change and a quiet ending:[20]

20.Marrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 63.

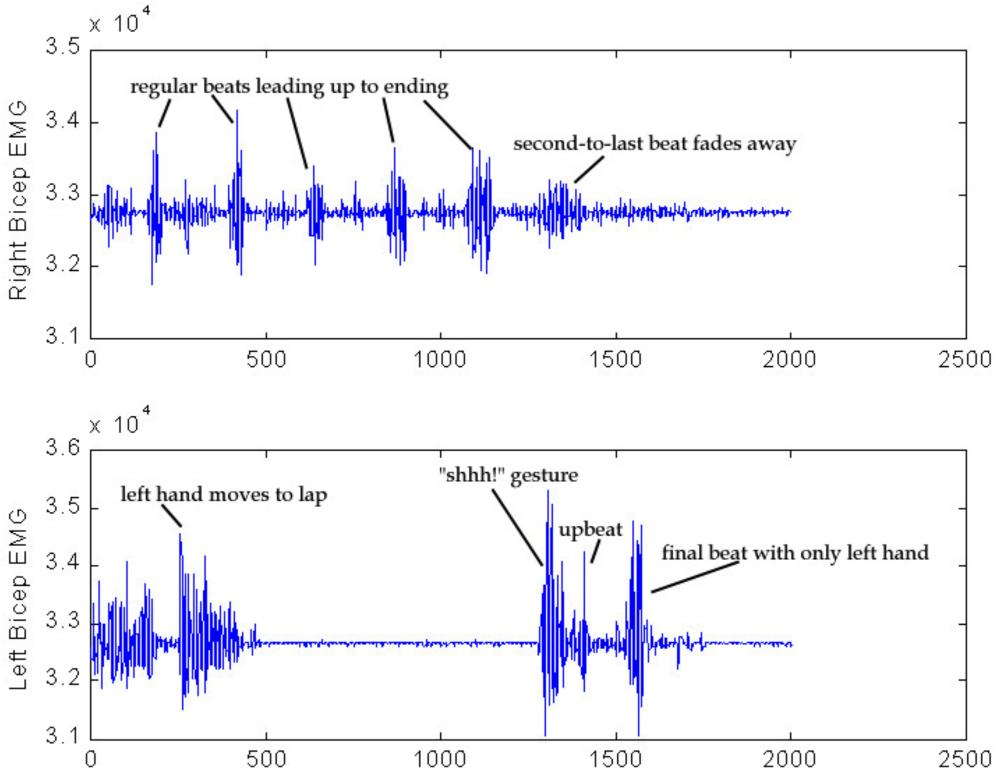

Fig. 3. Use of the left hand to indicate drastic change in volume.

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

In another example, P1 used his left arm to indicate alternations in groupings within the orchestra; these alternations were written in the score by the composer, and one of the functions of conducting is to exaggerate the structural elements in the score so as to create interesting contrast and character. In the opening theme of the Allegro movement of Tchaikovsky's Symphony No. 6, the strings introduce the main theme.

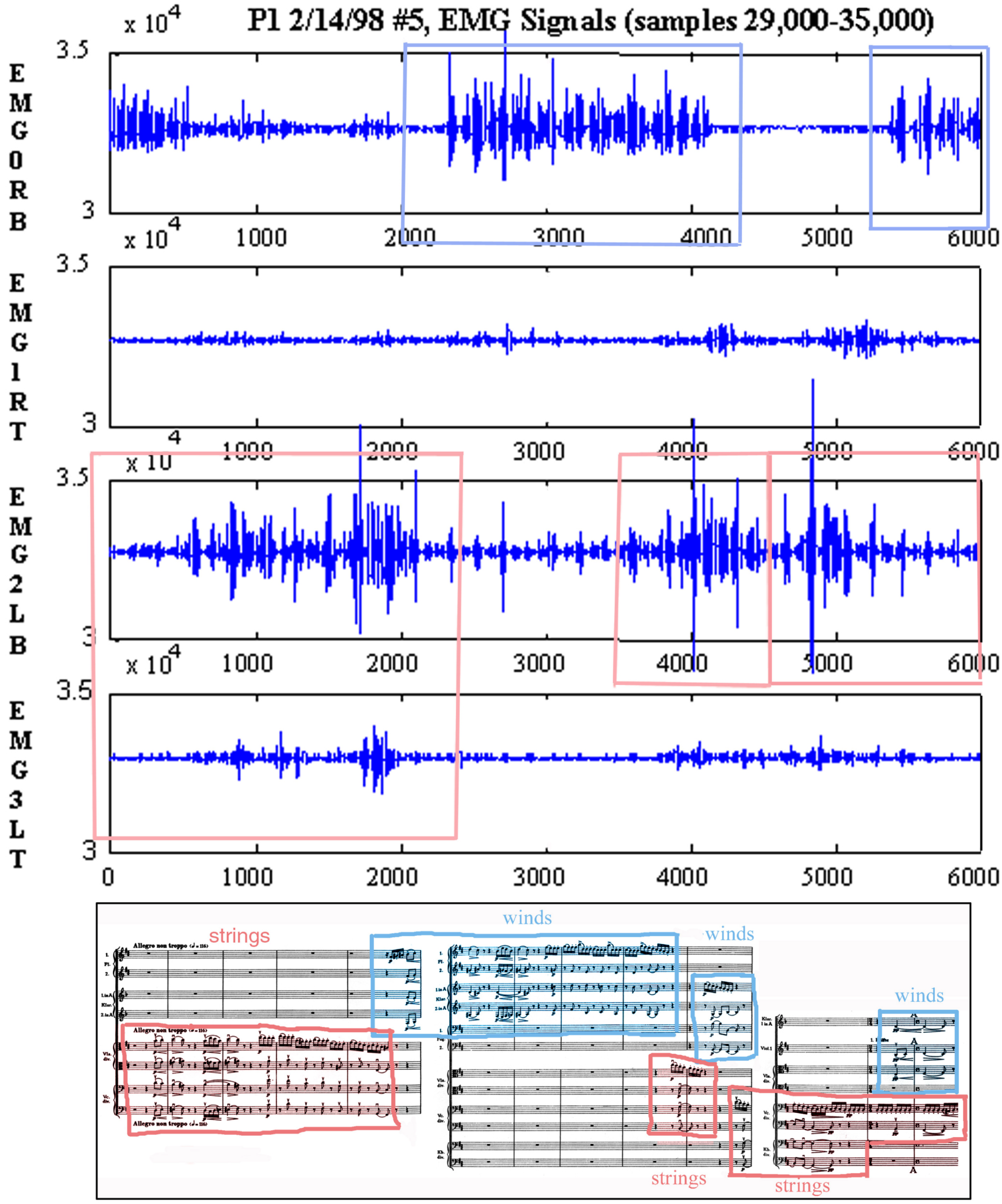

Fig. 4. Alternation between hands to indicate orchestral groupings.

In the second statement, the winds take over and repeat with slight variation what the lower strings just did. In order to show this, P1 suddenly minimizes the beats in his right hand and gives most of the gestures with the left hand. Figure 4, below, shows how P1 indicates the different orchestration groupings by switching arms. Note that the second time the winds come in, he does not give a separate indication – perhaps this can be explained by the quickness of the entrance or the dynamic marking of piano:

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

2. The flatlining effect

One of the most important functions of a conductor is to cue musicians to begin playing, particularly if they have waited silently for a long time; if the cue is not clear, they might not start playing in the right place. I found that our subjects would frequently withdraw gestural information suddenly before giving a cue. That is, their EMG signals (particularly from the left bicep) often went to nearly zero before signaling the onset of an important event or entrance. Such an event happened in P1's session at bar number 32 of the *Dance* movement in Prokofiev's *Romeo and Juliet Suite*; many of the woodwinds need to play after many measures of silence. Leading up to this event, P1 used his left hand normally, and then, two measures before the wind entrance, stopped using it completely. Then, just in time for the cue, he gave a big pickup and downbeat with the left arm. In repetitions of the same passage, the same action is repeated. This is demonstrated in figure 5, from P1's left biceps signal:[21]

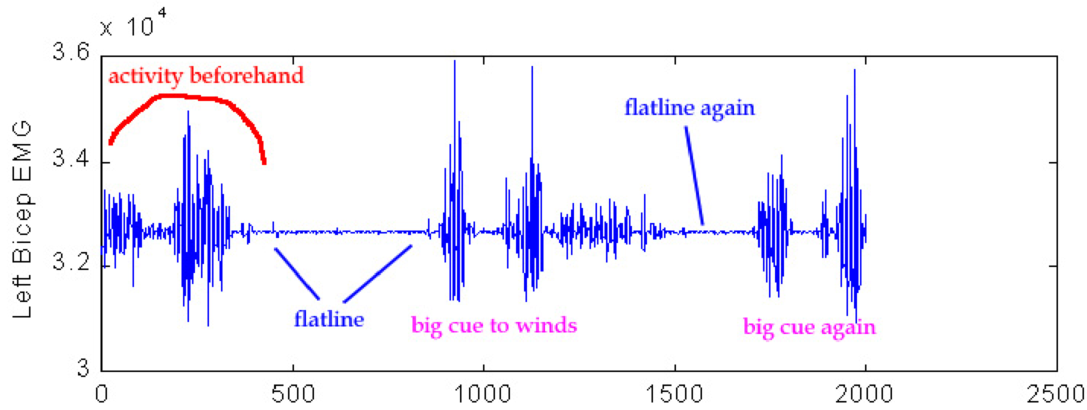

Fig. 5. The characteristic *flatline* in the left bicep before a major event

A reasonable hypothesis for why this "flatline" occurs could be that the sudden lack of information is eye-catching for the musicians, and requires minimal effort from the conductor. The quick change between information-carrying and non-information-carrying states could be an efficient way of providing an extra cue ahead of time for the musicians.

In a second example, P1 demonstrated this flatlining feature in his left bicep at the beginning of an extreme crescendo[22] in the string section. During this passage, corresponding to measures 3-7 after A in the last movement of Tchaikovsky's Symphony no. 6, P1's right biceps continued to beat regularly.

21.Marrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 63.

22.An Italian musical term that indicates increasing loudness.

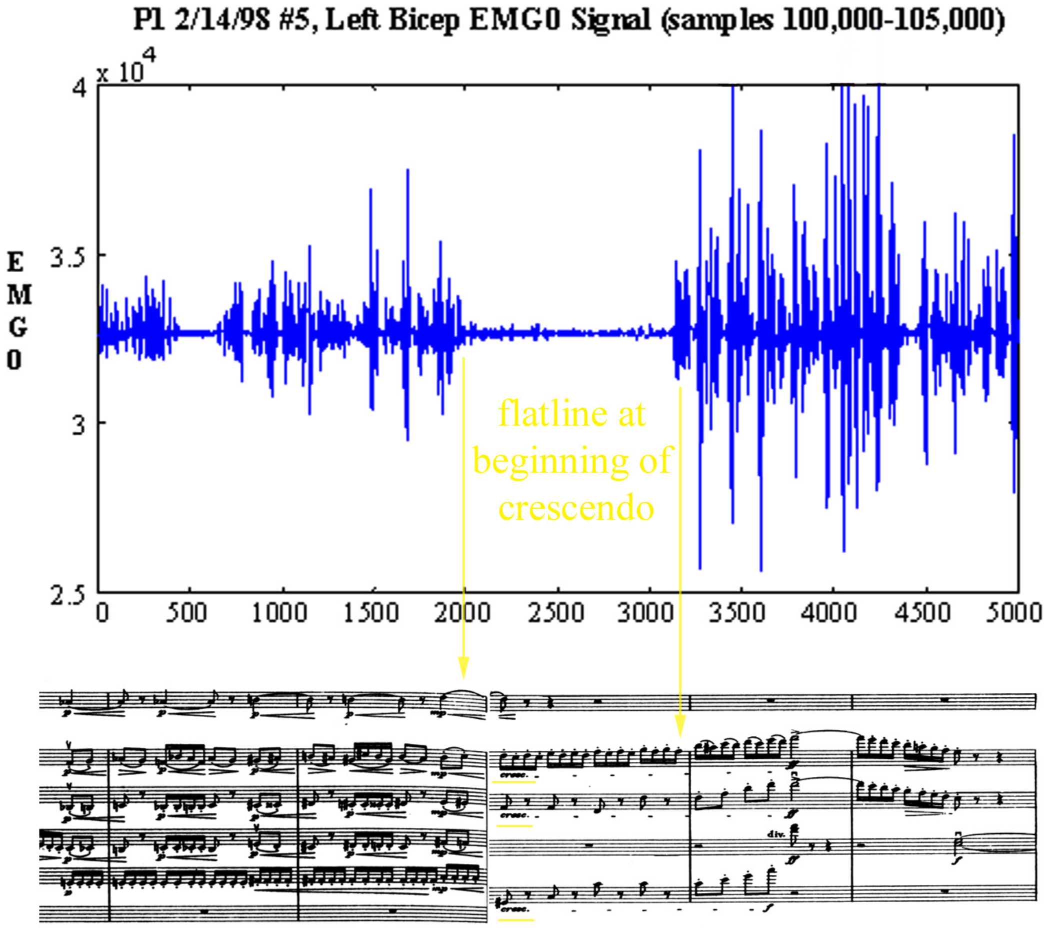

Fig. 6. Characteristic *flatline* phenomenon right at the beginning of a large crescendo

In an example from subject P3 I found two successive flatlines, flanking a gradual decrescendo.[23] These events were interspersed by four repetitions of a theme. These repetitions alternate in emphasis, switching twice from medium loudness (*mezzo forte*, *mf*) to extra loudness (*fortissimo*, *ff*). The segment begins with the second theme of Sousa's *Washington Post March*, followed by an answer in a complementary theme (conducted softly in the right bicep while flatlining in the left) followed by a second statement of it (conducted forte in both arms). This is followed by a bridge (conducted piano with less motion in both arms), leading to a repetition of this whole section. Leading up to the fourth repetition, perhaps for emphasis, there is a longer flatline section and a much higher-amplitude EMG signal.

23.An Italian musical term for decreasing loudness.

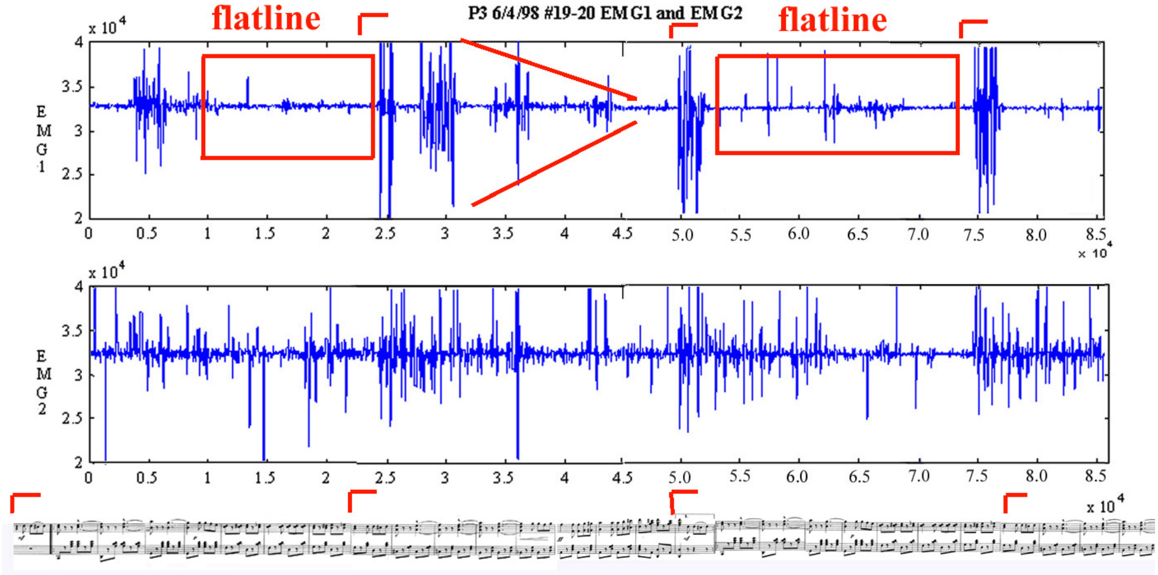

Fig. 7. Four consecutive statements of a theme, interspersed by flatlines and a decrescendo

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

It is possible that the *flatlining* phenomena in these examples from conductors are reflective of similar phenomena in other areas outside of human musical behavior. For example, snakes coil before striking, cats pause before pouncing, and windspeeds and barometric pressure decrease right before the arrival of storms.

3. The direct, quasi-linear correlation between muscle tension and dynamic intensity

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

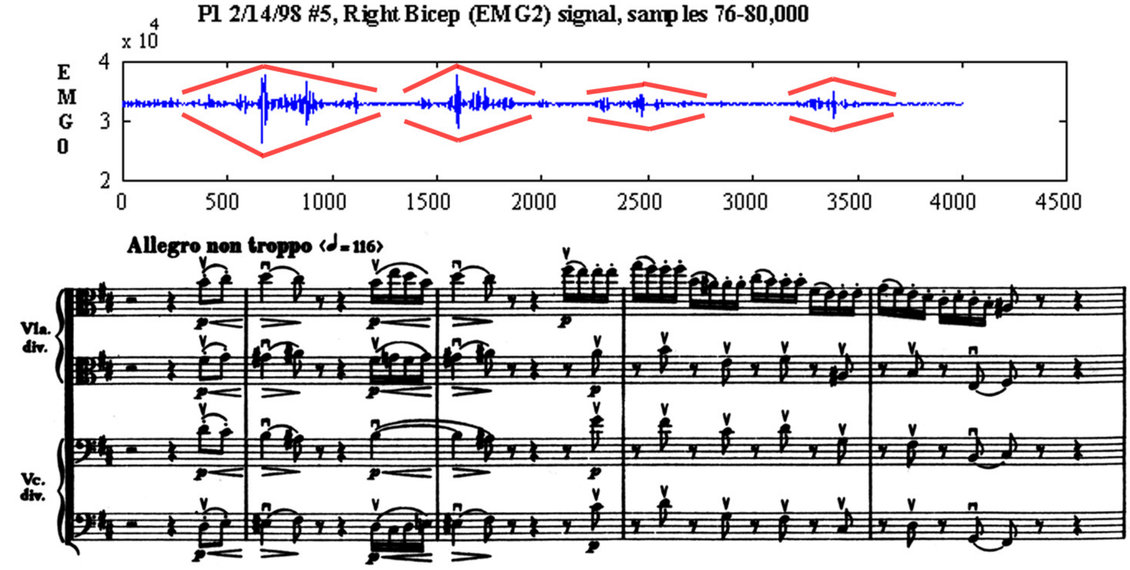

Fig. 8. P1's crescendo-decrescendo pairs in the right biceps

In a second example approximately thirty seconds later, P1 gives a large-scale crescendo-decrescendo that takes nine measures to complete. Again, the relative scale between the softest and loudest portions appears to be roughly linear. In figure 9, below, P1's EMG signals scale from pp up to ff and back down to pp; this corresponds to 1 before A to 12 after A in the fourth movement of Tchaikovsky's Symphony no. 6. All four EMGs are shown – EMG0 is his right biceps, EMG1 is his right triceps, EMG2 is his left biceps, and EMG3 is his left triceps.

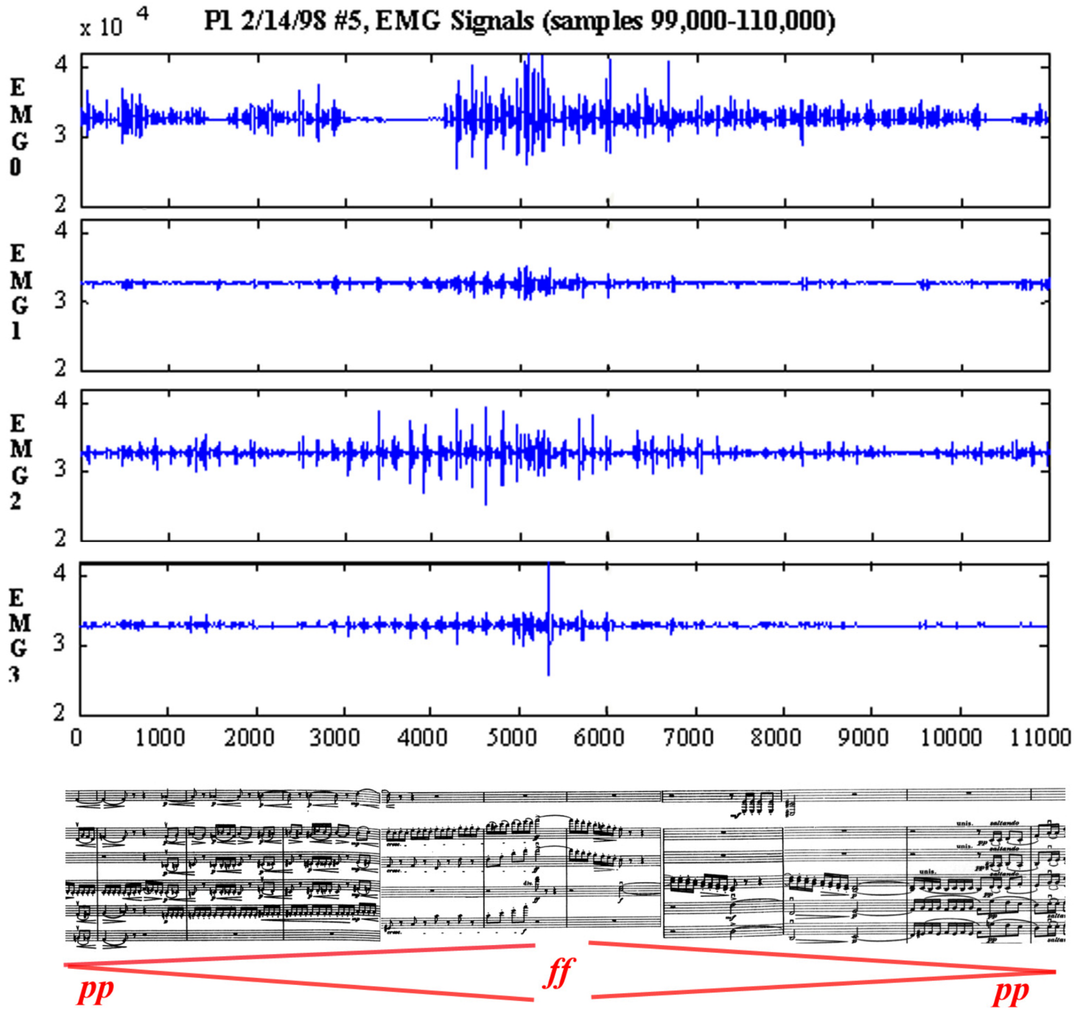

Fig. 9. EMG amplitude expansion and contraction during an extended crescendo-decrescendo

In a third example, subject P3 makes a graduated decrescendo and crescendo with his left hand during a segment from the *Love Theme* from *Titanic*. Figure 10, below, shows his left biceps' signal during this segment, demonstrating the extreme difference between various levels of tension in this muscle.

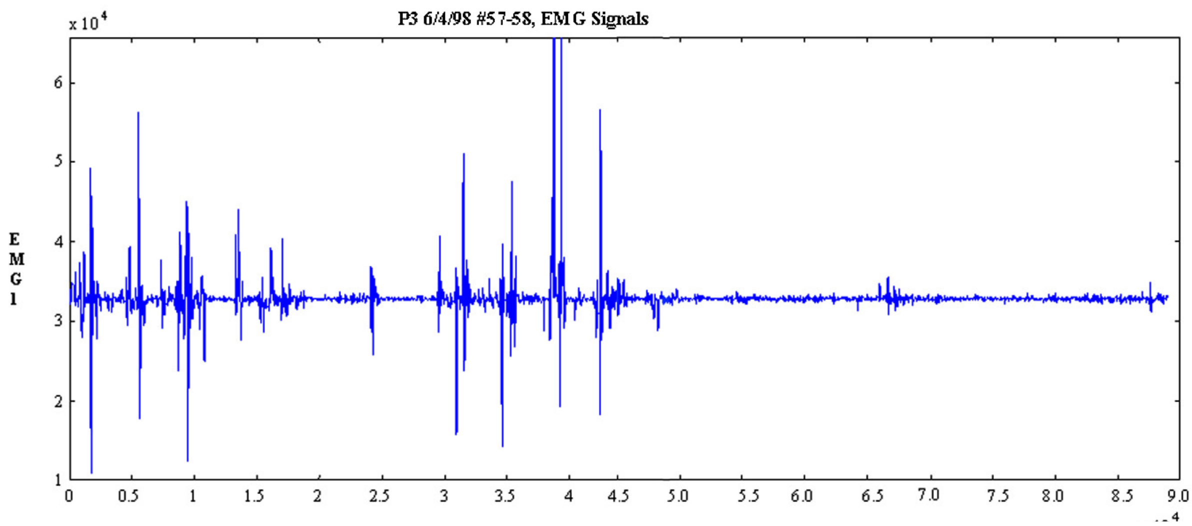

Fig. 10. P3's left biceps signal during the *Love Theme* from *Titanic* showing graduated crescendo-decrescendo values

4. Predictive indications

In addition, there appears to be a kind of 'predictive' phenomenon, whereby our conductor subjects indicated specific events on the beats directly preceding the intended ones. This is often discussed in the literature on conducting technique, as evidenced by a sentence in one of the most influential treatises on conducting: "in a sense, all conducting is preparation – indicating in advance what is to happen."[24] Also, in another important text, the author urges conductors to "bring the left hand into play one beat before the cue-beat and make a rhythmic preparatory beat leading to the cue. These are both good for improving your 'timing.'"[25] Despite numerous mentions of this predictive phenomenon in instructive and pedagogical documents, this phenomenon has not previously been shown to be true with any quantitative methods. Figure 11, below, shows a segment of the score to Prokofiev's *Dance* movement (from *Romeo and Juliet*) with the accents highlighted and aligned with the accents given by our subject.[26] As indicated, the accents are given one beat ahead of the intended beat:

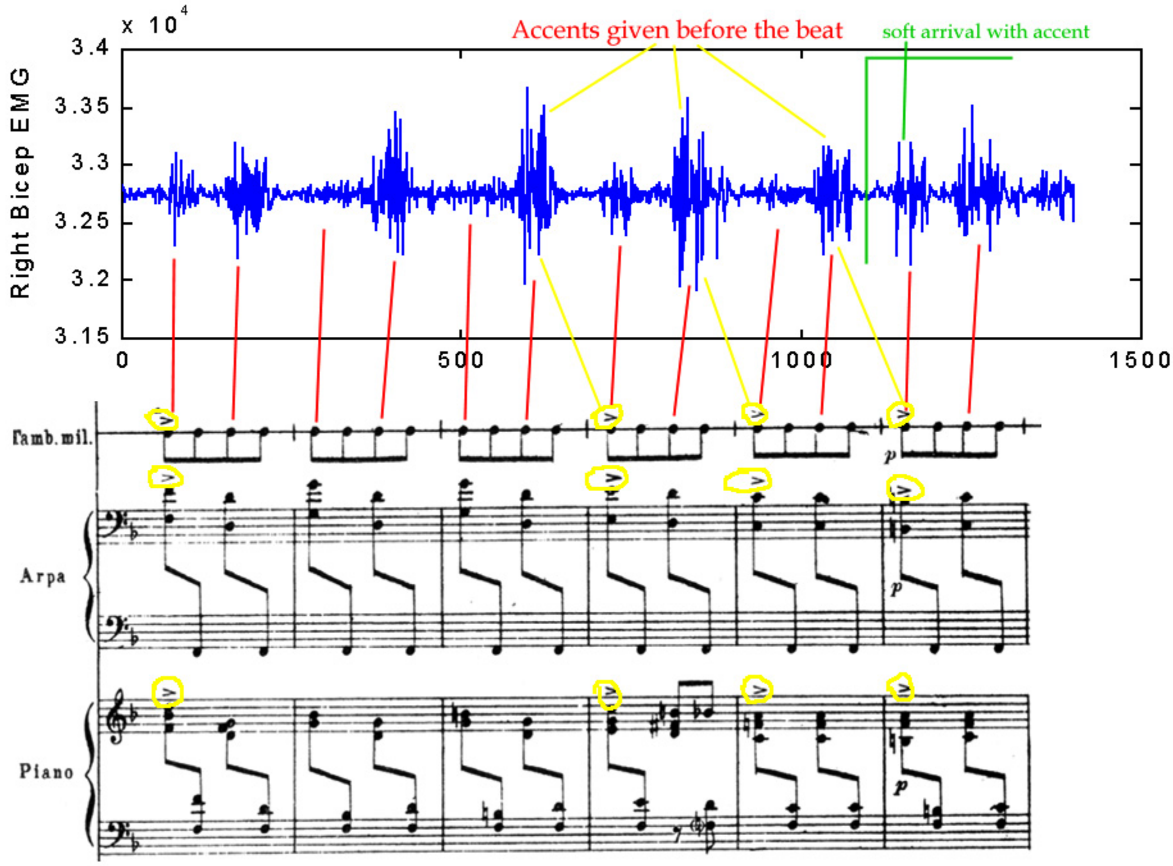

Fig. 11. 'Predictive' accents and their relation to the score

The red lines show how the EMG signals line up with the beats in the score, while the yellow lines show the relationship between the conducted accent and the accent where it is to be played. The separation line around sample 1100 represents the barrier in-between the first, loud, section, and the second section, which is to be played quietly. The reduced amplitude of the EMG signal right before the separation line could indicate an anticipation of the new loudness level. Also, the existence of high-amplitude EMG signals on non- accented notes in this passage cannot be accounted for musically; perhaps they are due to conductor error. Alternately, they might be accounted for from the perspective of information theory, that once a pattern has been established it does not have to be indicated at every point.

24.Max Rudolf, *The Grammar of Conducting*, page 317.

25.Elizabeth A.H. Greene, *The Modern Conductor*, p. 90.

26.Marrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 64.

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

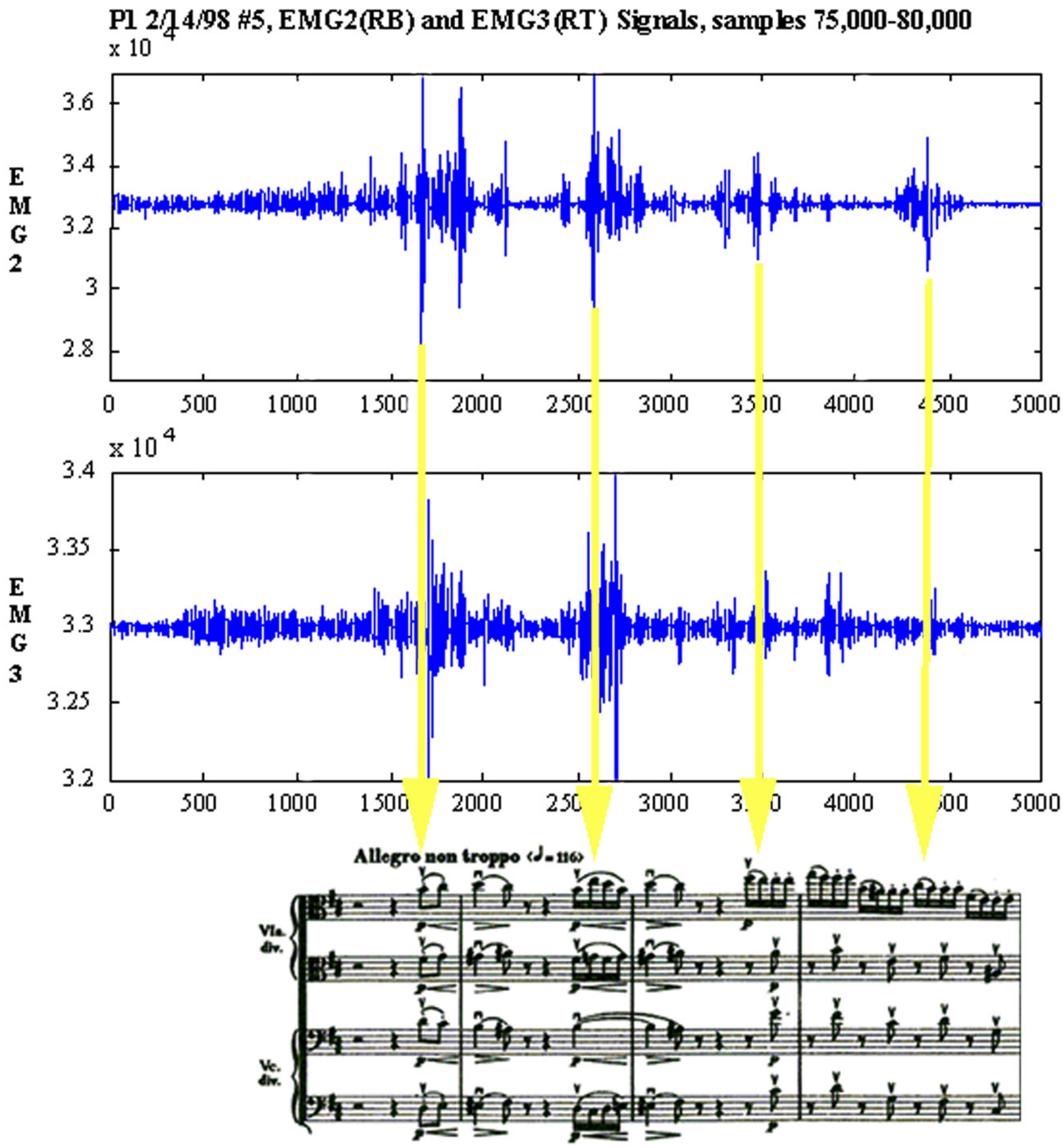

Fig. 12. P1's predictive indications of emphasis

5. Repetitive signals minimized until new information appears

Another consistent feature among our subjects was that after a few repetitions of something, the signals would be minimized until new information needed to be conveyed. Professionals tended to reduce the size and velocity of their gestures during segments where there was no change; when new musical information approached, then their gestures became bigger and faster.

For example, after the opening of Sousa's *Presidential Polonaise*, P3's gestures decreased in size because the tempo and musical character did not change for a few bars. This successive lack of emphasis in repetitive, static sections is significant because it is a case in which the relationship between the signal and its intended meaning is nonlinear. That is, since no change is required, the slow decrease in amplitude of gestures allows the conductor to slowly reduce his effort without causing the musicians to similarly decrease the amplitude of their efforts.

Figure 13, below, demonstrates this phenomenon.

This phenomenon suggests that conductors operate according to an efficiency principle. When new information is not needed by the orchestra, the overall intensity level of all gestures is reduced. This point is supported by the conducting pedagogue Max Rudolf, who wrote, "you should not use more motion than you need in conducting."[27]

Perhaps the prevalence of this feature indicates experience or expertise on the part of the conductor.

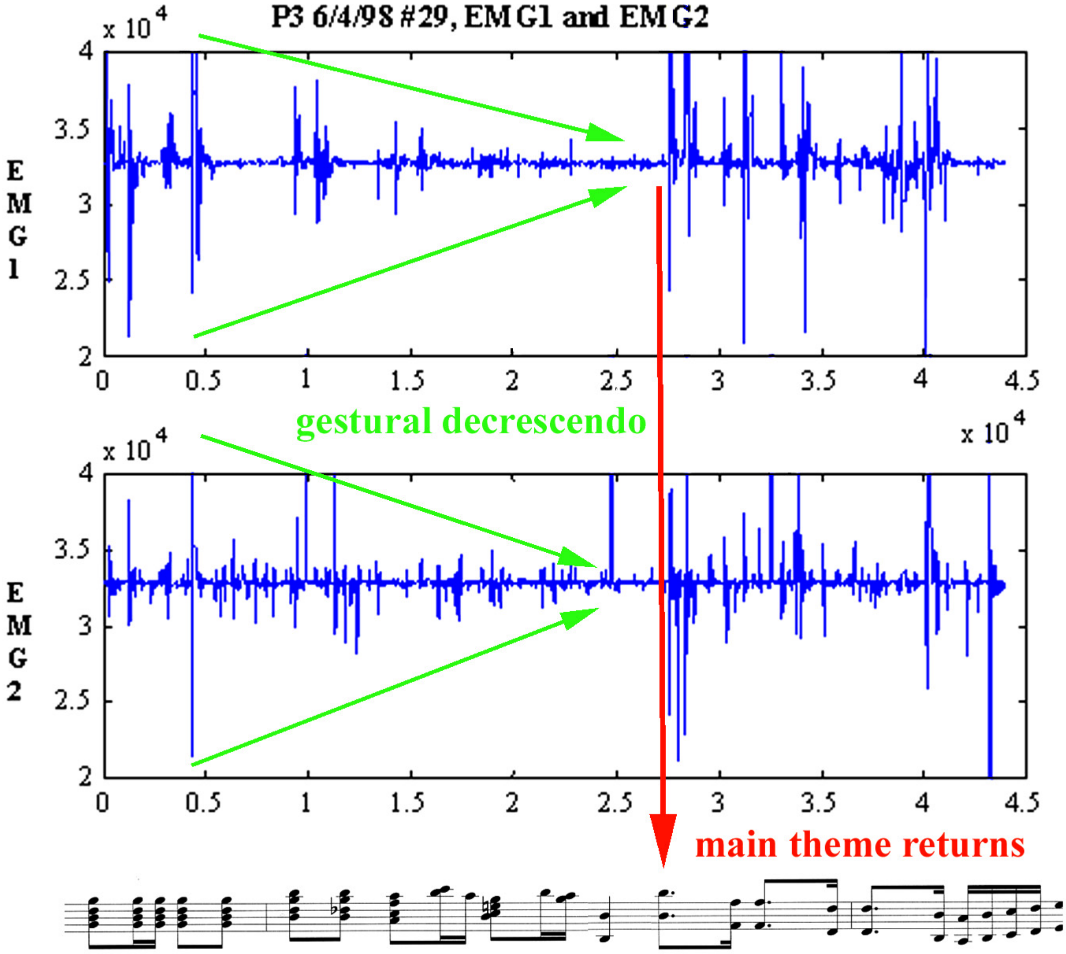

Fig. 13. Gradual diminution of EMG amplitude during repetitive sections

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

Professionals showed fundamental differences in the way they made information-carrying gestures vs. non-information carrying gestures (students less so). A third feature I discovered in the EMG data is that the signals generated by the action of turning pages are inherently different in character from the signals generated by actions that are intended to convey musical information. That is, it seems as if page turns are done in such a way as to purposefully not attract attention or convey musical information. An example page turn from subject P1 is shown below in figure 14:[28]

27.Max Rudolf, *The Grammar of Conducting*, p. 314.

28.Marrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 63.

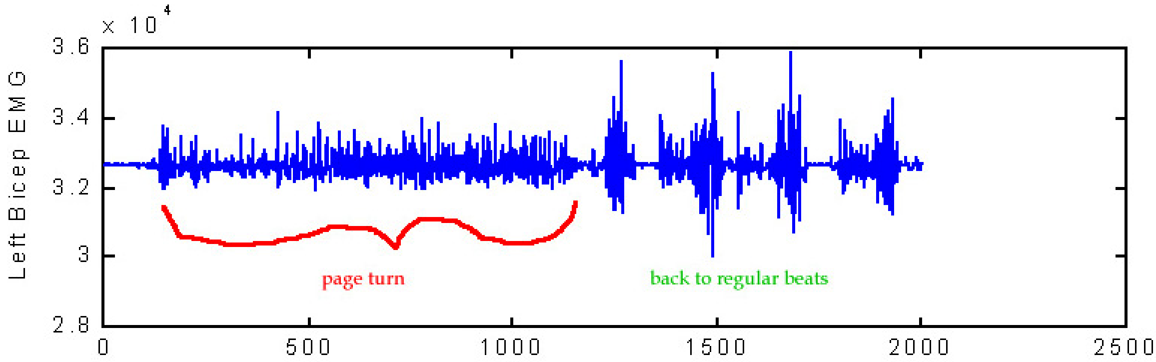

Fig. 14. Difference between page turn gestures and information-carrying gestures

A corollary to this is that the professionals tended to generate far less muscular force for noninformative events than the students do.

For example, P3 typically used almost no force when turning a page.

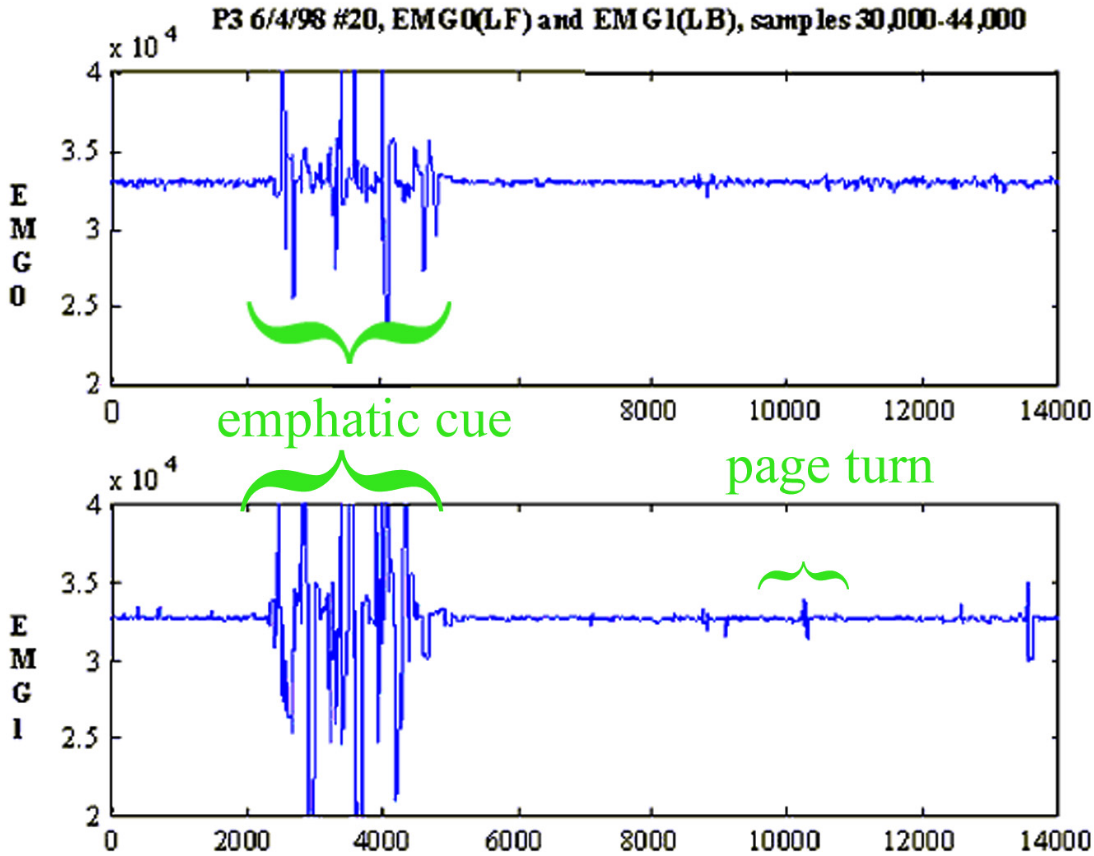

Fig. 15. P3's page turn gesture compared with a strong musical cue

As in figure 15, above, the EMG0 (left forearm) and EMG1 (left biceps) signals show the relative difference between an information-bearing and non-information-bearing gesture. The first feature is an emphasis cue in the left arm; the following feature (samples 40-44,000) is the page turn. In the page turn gesture, the arm moved to perform the action, but without force or emphasis. This lack of muscle tension during the gesture is significant; perhaps this is because musicians are sensitive on a subconscious level to muscle-tension cues and therefore are trained to ignore gestures that do not have tension in them.

7. Frequency of unnecessary actions decreases with experience

As a corollary to the above point, I found that students tended to do unnecessary actions more often than professionals, and with greater muscle tension. These actions included movements such as pushing and pulling at the music stand, scratching, touching their hair, etc. These unnecessary gestures possibly provide a cue to nervousness as well as inexperience; some unnecessary actions are certainly natural and normal, but the student subjects tend to perform them more often than the professionals. In the cases where the student subjects adjusted their stands, the movement upwards or downwards was frequently accompanied by a gesture to push the stand away. One explanation for this is that people tend to push things away when they feel aversive or repulsed, and perhaps the stand is unconsciously used to demonstrate that feeling.

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

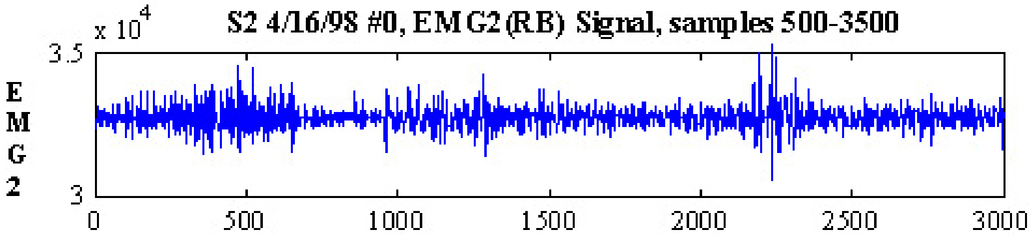

Fig. 16. S2 begins with three unnecessary actions

In another example from a student subject, S3 pulls up her stand forcefully after several minutes of conducting. There is no apparent reason for this action and no need to use such force for the action; her right biceps EMG signal from that event is shown below in figure 17:

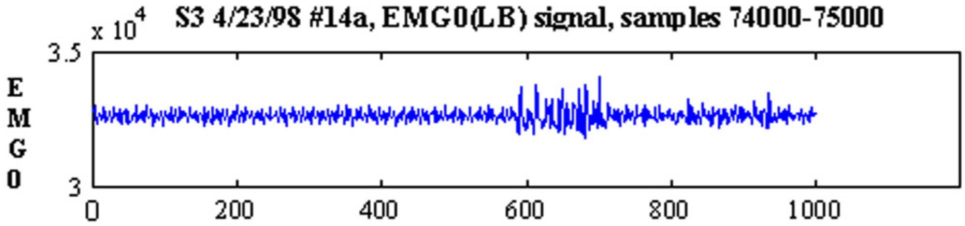

Fig. 17. S3 pulls at the music stand

The student subjects also exhibited many other unnecessary actions such as twitches, apparently due to nervousness and lack of control. Compared with the amplitudes of signal-carrying gestures, the amplitudes of these are high and could interfere with musicians' perception. S3 tended to often place her left hand on her head or touch her hair with her left arm, as shown below in figure 18:

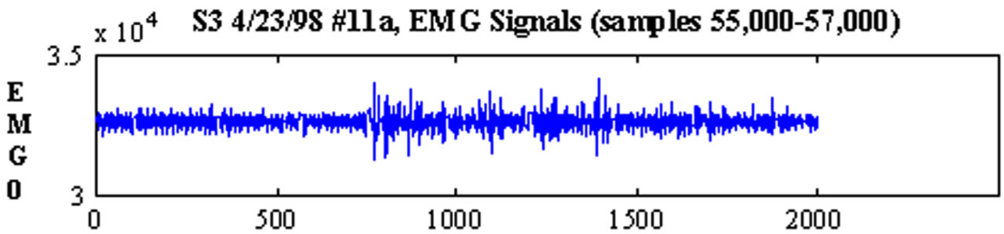

Fig. 18. Subject S3 touches her head and generates muscle tension signal

In another example, S1 scratches his back with right hand (EMG2) and shakes left hand (EMG 0). These signals are relatively much larger than the beats that preceded them; the beats are located in samples 0-300, the back-scratching signal is in samples 400-900 of EMG2, and the hand-shake is in samples 800-1200 of EMG1. These signals are shown in figure 19, below:

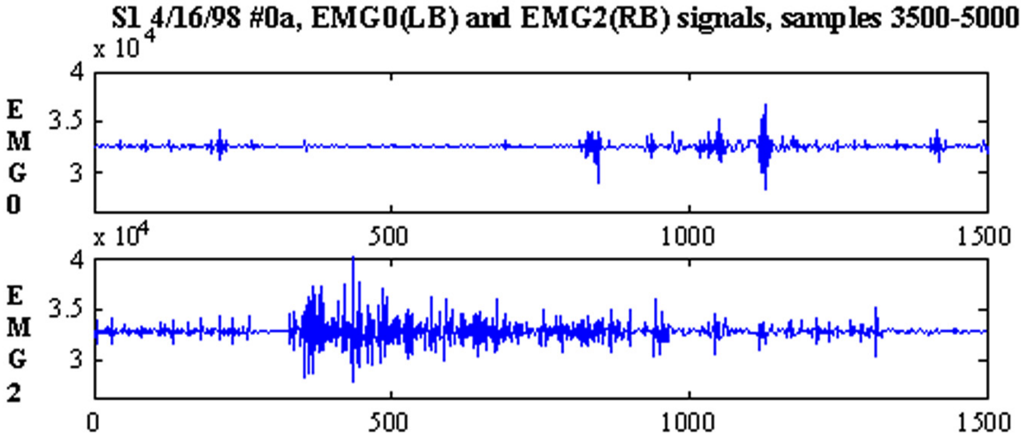

Fig. 19. S1's left and right biceps signals, showing small beats followed by scratching his back and shaking his hand

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

8. Clarity of signal during slow, legato passages correlates with experience

In general, the students' EMG signals lost clarity as the tempo slowed or when the character of the music became more sustained.

The definition of their beat signals was almost entirely lost in *legato*[29] or slow music. In faster, more accented music, the beat envelopes became much clearer. For S1, legato gestures in the right biceps gradually became mushy after a few accented beats, beginning at sample 1200, below:

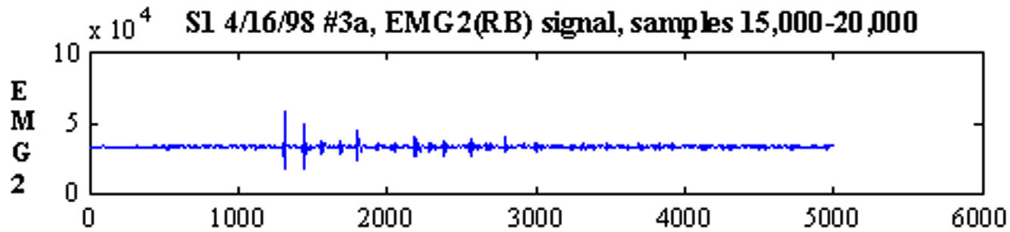

Fig. 20. Gradual decrease in clarity of beat signal during legato passage

29.An Italian musical term meaning smooth, continuous articulation with little separation between notes.

Similarly, for subject S2, the clarity of beats decreased in slow music, as demonstrated below in figure 21:

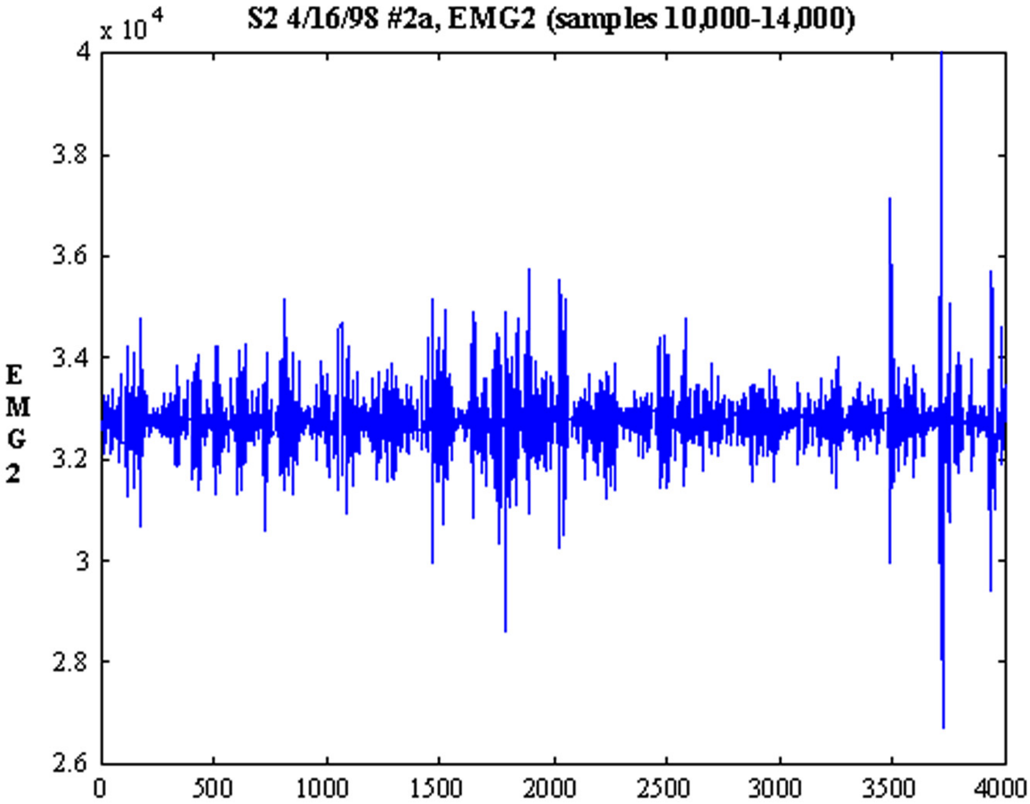

Fig. 21. Indistinct beats during a slow passage from subject S2

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

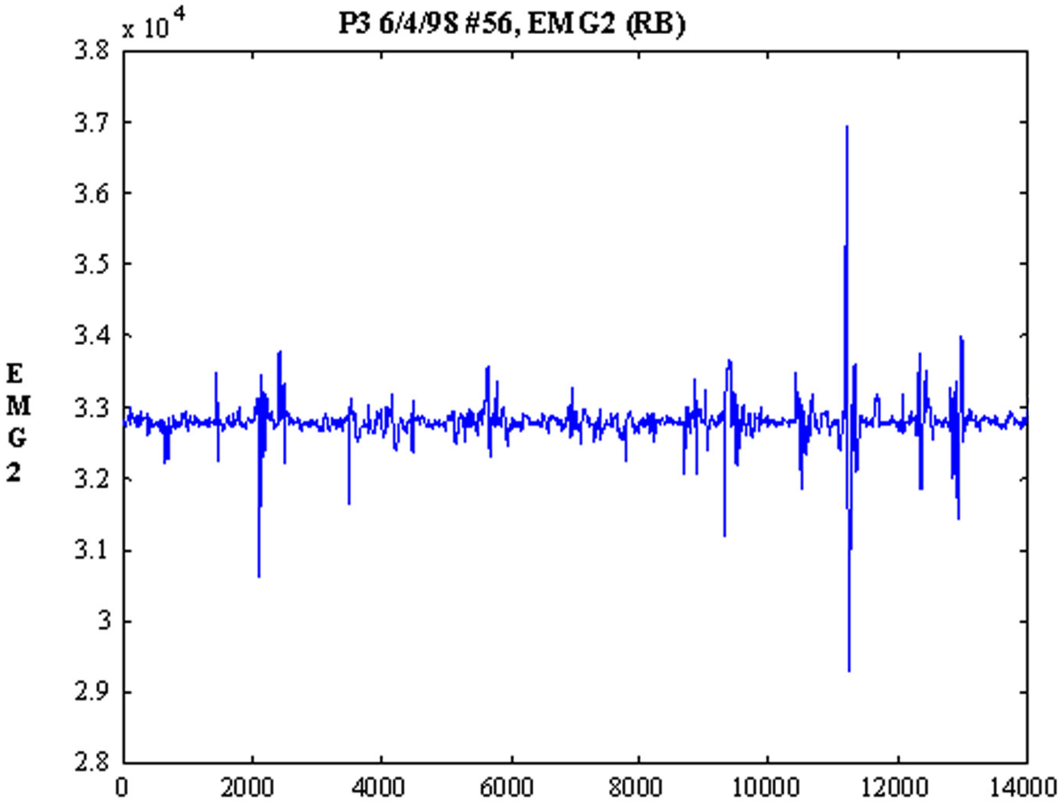

Fig. 22. P3's right biceps signal during a legato passage

9. Division of labor between biceps, triceps, and forearm

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

For example, in the *Love Theme* from the movie *Titanic* (the only piece in P3's program which was uniformly legato in style), there were several places where the amplitude of his right forearm EMG signal was larger than that of his right biceps. Also, there were several places where the forearm gave different information from what the biceps did; this was extremely unusual. At the very beginning of the piece, the biceps signal was almost nonexistent, and the largest signal came from the right forearm, the muscle that most significantly influenced the use of the baton. This example is given below in figure 23; at the end of this segment, the forearm gives large beats that usher in the main theme of the piece:

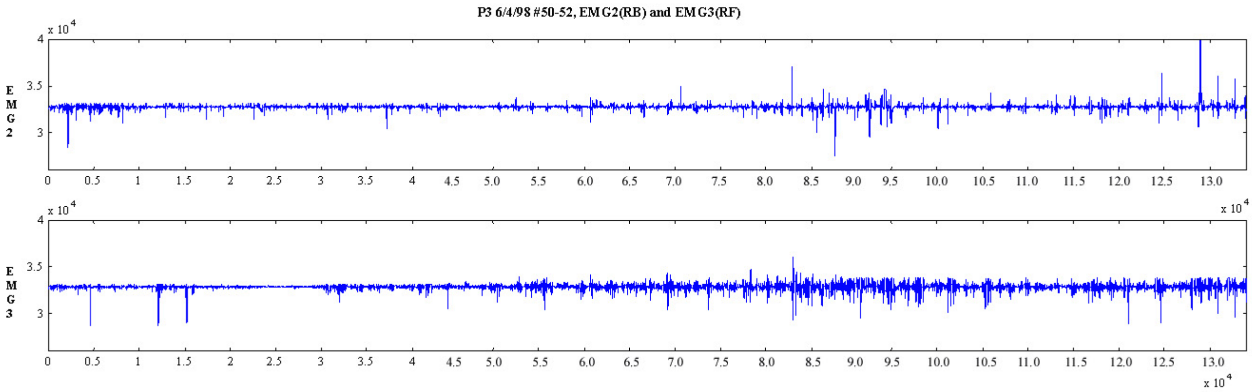

Fig. 23. P3's forearm signals eclipse the biceps signals during legato passage

During this example, the relative size of the gesture remains large (as evidenced in the videotape from the segment), but the tension in the muscles is not present. This phenomenon is observable visually and palpable viscerally, but not easily quantifiable.

From these results, it seems that a forearm EMG measure is more useful than a triceps measurement; the biceps signal, however, is essential. It's not clear if more than two EMG sensors per arm would yield useful results for larger, conducting-style arm motions.

10. Rate encoding

Rate encoding is the modulation of the frequency of repetitive events in order to imply amplitude changes in another domain; it is similar to frequency modulation, but instead of reflecting the frequency of a sinusoid, it is the frequency of some specified event.[30] Conductors use rate encoding (often in the form of subdivisions within the beat pattern) to specify intensity or dynamic changes. For example, in a segment from subject P1, he starts the crescendo by increasing the amplitude of EMG tension with his right arm, then adds left arm, and then doubles the frequency of his beats for the last two bars. That is, instead of giving beats in a meter of one for every two quarter notes, he gave beats in a meter of two. The example of rate encoding in figure 25 corresponds with measures 5-8 after A in Tchaikovsky's score to the Symphony no. 6;[31] P1's EMG0 signal shows the left biceps signal and EMG2 signal reflects the right biceps.

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

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

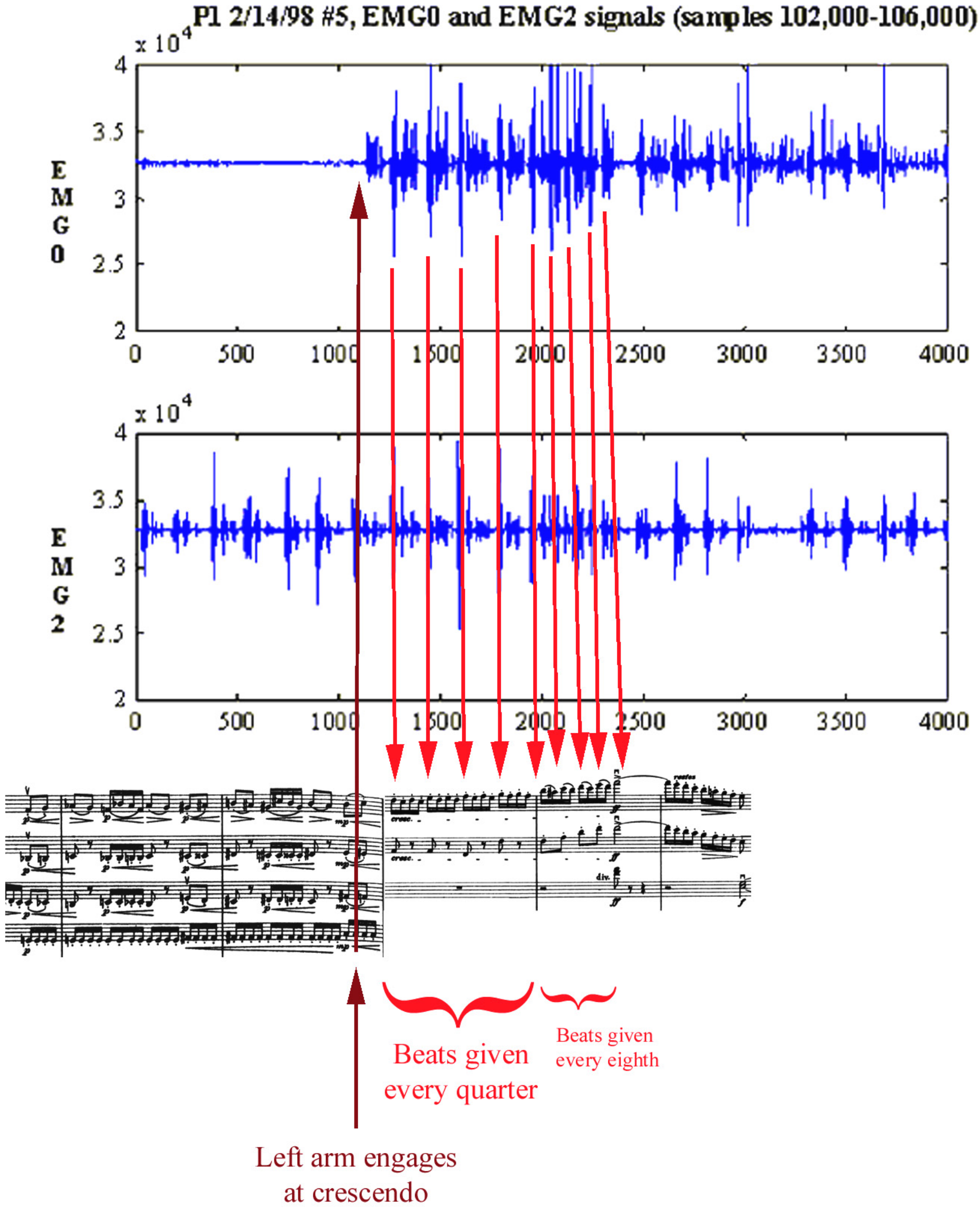

Fig. 24. Rate encoding evidenced in P1's beat signals compared with the score

11. The link between respiration and phrasing

The data demonstrated numerous examples where the subjects modulated their natural respiration cycles in order to reflect phrasing and character in the music. This modulation may or may not have been conscious and purposeful, and may also have been influenced by the motions of their arms, but nonetheless seems significant and highly correlated with the expressive qualities in the music. For example, P1's breathing correlated more closely with the phrasing in the music than that of any other subject. In one musical section, P1's respiration cycles matched the metrical cycles of the music; when the meter changed, so did his breathing patterns.

Secondly, his respiration signal often increased in anticipation of a downbeat and sharply decreased right afterward. This might have been the result of the compression of the ribcage in the execution of the beat, but could also be an intentionally expressive phenomenon.

For example, it is considered a standard practice among conductors to breathe in at upbeats and breathe out at downbeats, regulating their flow of air relative to the speed and volume of the music. [32]

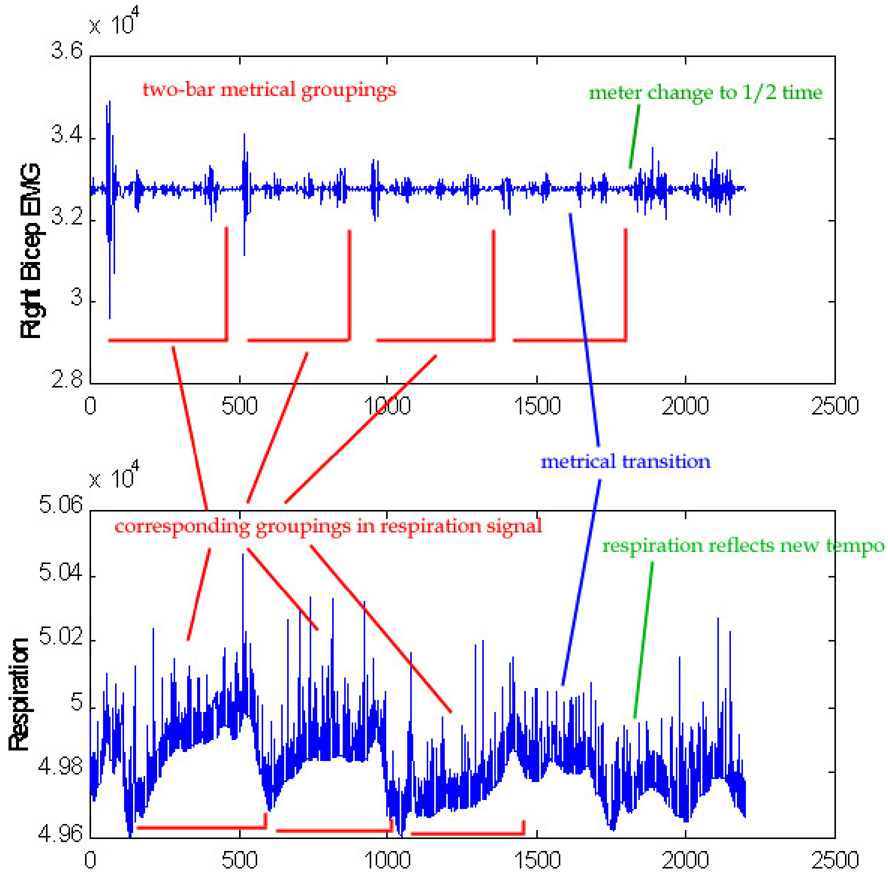

Fig. 25. P1's respiration signal compared with his right bicep signal to show correlation

Similarly, P3's breathing showed a very tight correlation with his beat gestures – during loud, active, regular passages (such as during Sousa's *Washington Post March*, that began his program), the respiration signal seemed to correspond directly with the beats. It is not clear, however, what the causes and responses are, and whether this is a case of arm motion completely dictating the movement of the ribcage, or conscious breathing with the beats.

12. Large GSR peaks at the beginning of every piece

In addition, across most subjects, there tended to be a large increase in the GSR signal before the beginning of each piece. It is possible that this is this evidence of motion artifact from the use of the left arm, but it is also possible that it could indicate a significant trend in skin conductance changes.

For example, subject P1's GSR baseline increased markedly right before beginning a very expressive segment with the orchestra; his signal level remained high throughout the segment. Then, after stopping the orchestra, his GSR signal decreased back to earlier levels. In figure 27, below, P1 signals the orchestra to start playing at sample 1700, and signals it to stop at approximately sample 5000:

32.eMarrin, Teresa and Rosalind Picard. "Analysis of Affective Musical Expression with the Conductor's Jacket." *Proceedings of the XII Colloquium on Musical Informatics*, Gorizia, Italy, 1998, p. 64.

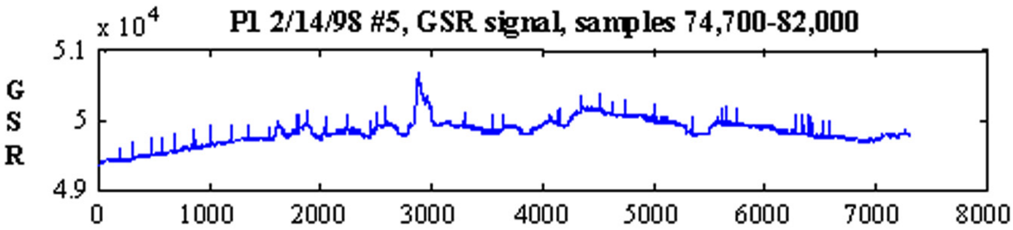

Fig. 26. P1’s GSR signal at the start of a rehearsal segment

13. GSR baseline variance as a strong indicator of experience

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

Figure 28, below, illustrates the activity of P3’s signal during the course of the segment:

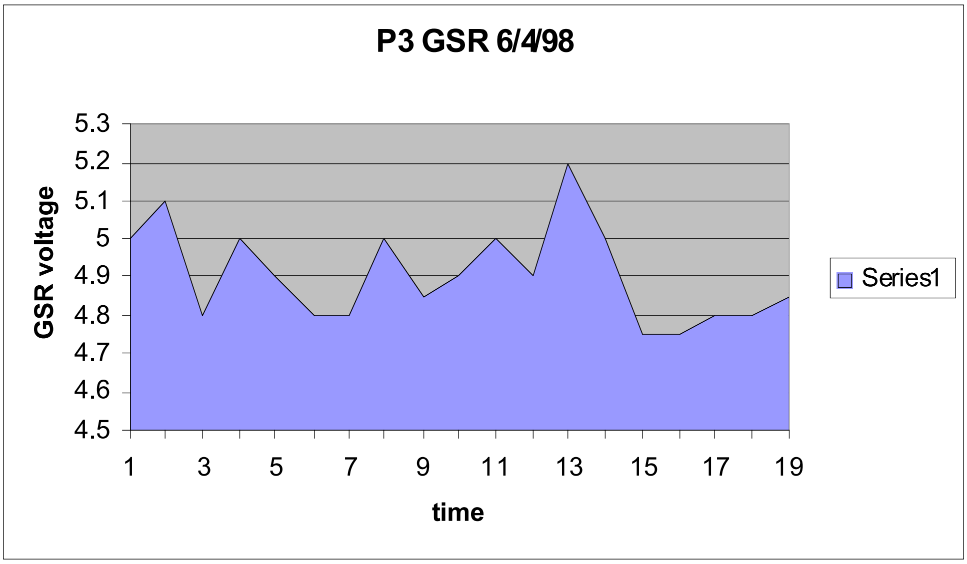

Fig. 27. Trajectory of P3’s skin conductance during the course of a concert

14. Temperature baselines

In the case of temperature baselines, the comparisons could not be made as clearly as they could with GSR. For example, S1’s temperature had a .01 volt swing[33], S2’s temperature had a .02 volt swing[34], S3’s temperature had a .03 volt swing[35], and P1’s temperature had a .02 volt swing[36]. P3’s temperature was the

---

33.It began at 5.47 and ended at 5.48 after almost 5 minutes.

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

most active: it had a .11 volt swing. This might be explained by the fact that the data was taken during a live performance. P3's temperature trajectory is given, below, in figure 29; it reflects his temperature signal during a forty-minute segment in the third and last act in the concert.

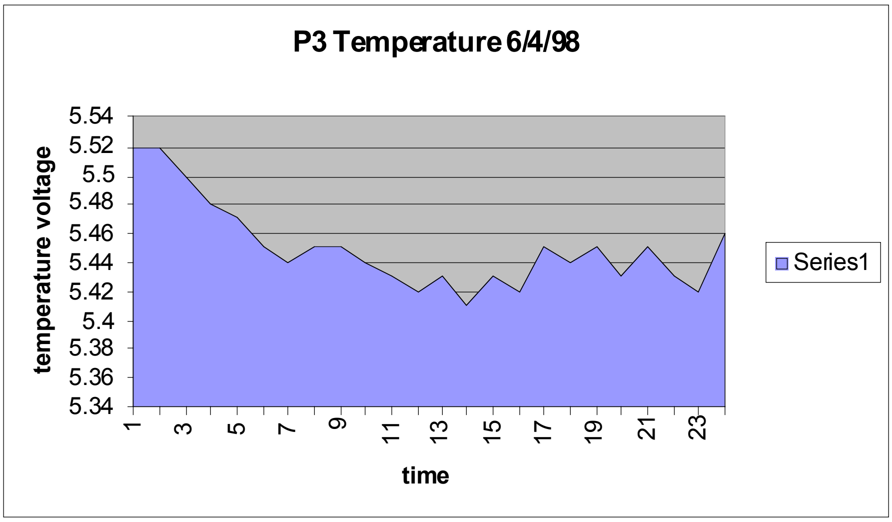

Fig. 28. P3's temperature profile during the last segment of a concert

During the entire segment, P3's temperature slowly decreases with a small upward feature at the end. At the beginning of the third act, his signal begins at 5.52 volts. During the first piece, his signal quickly decreases to 5.48. During the second piece, the temperature signal starts at 5.47 and slowly decreases to 5.45. During a short interval, his signal decreases slightly to 5.44, then increases slightly to 5.45 while talking. During the third piece, his signal starts at 5.45 and slowly decreases to end at 5.44. During the following interval, it decreases quickly. Then, during the fourth piece, it begins at 5.42 and ends at 5.43. During the next interval, the signal decreases down to 5.41 and increases slightly. The fifth piece began at 5.43, decreased to 5.42, and slowly increased back up to 5.45 by the end. The sixth piece began at 5.44 and remained constant until its ending at 5.45. The seventh piece began at 5.43, increased up to 5.45, and decreased to 5.43. The eighth and final piece began at 5.42, decreased slightly, and then ended at 5.46 volts. One hypothesis about the greater activity of P3's temperature relative to the other subjects is that the stressfulness of the concert situation makes temperature more variable; the five other subjects had a much more static temperature value.

Conclusions

These preliminary results of the Conductor's Jacket project point to the strong possibility for a rich new area of study. However, much remains to be done. Now that fourteen significant results have been found in the data, many more experiments should be conducted to support them with statistical methods. Secondly, algorithms will need to be written to appropriately segment, filter, recognize, and characterize these features; they will make the data more usable by exposing its underlying structure.[37] Segmentation will be necessary in order to pick the areas where the data is richest, such as conducting vs. non-conducting, informative vs. non-informative gestures, and beginnings and endings of pieces. Automatic recognition tasks will involve the implementation of feature detection systems using models with properties such as clustered weights, hidden markov processes, or hierarchical mixtures of experts. Finally, with these filters

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

36.It began at 3.35 volts, decreased to 3.34, increased gradually to 3.36, dipped down briefly, and returned to end at 3.36.

37.This work is underway and will be discussed in my forthcoming doctoral dissertation.

and recognition algorithms adapted to real-time, it will be possible to build systems to use these expressive features as control signals for a variety of live interactions, including live musical performances and multimedia artworks.

One phrase that comes up frequently in studies of musical expression is ‘musical intention.’ It is often used to describe the actions that a musician makes while performing which reflect their musical feelings and ideas (these can either be directly or secondarily related to the gestures required to produce the sound). Quantifying musical intention is a very difficult problem, but it might be achieved by separating intentionally expressive from intentionally unexpressive gestures. The preliminary results from the Conductor’s Jacket study indicate that intentionally expressive gestures are those for which there is a prominent EMG feature (i.e. where force is generated for emphasis, irrespective of the size and velocity of the physical gesture). Conversely, gestures that have large velocity and size but less EMG amplitude can be said to be intentionally unexpressive. As has been shown, conductor experience influences the validity of this claim; students often created large signals during gestures that had nothing to do with the musical content of what they were doing.

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

Future work in this area will necessarily focus on extending and improving upon the mappings that take gestural signals and convert them to music. These mappings must be intuitive and powerful enough to respond appropriately to the structure, quality, and character in the gestures. They must not only satisfy the performer’s needs for intuition and naturalness, but also make the gestural- musical relationships clear to the audience that is witnessing the performance. More complete systems might also include measures for eye gaze, facial expressions, gesture recognition, and self-report or commentary from the conductors themselves. A future study might also make use of the following axes of comparison, which were applied informally in the Conductor’s Jacket study. Performing a full statistical correlation on features that are found to consistently differ between contrasting pairs could yield very interesting results: 1. Within musical events 2. Within pieces – between different musical sections, between repeated sections or motifs 3. Between different pieces, particularly between ones of contrasting character, such as energetic vs. subdued (Sousa vs. *Titanic*) 4. Within similar musical styles 5. Between different musical styles 6. Across conductors 7. With regard to conductor experience and training 8. With regard to a conductor's opinions about what happened 9. Over the course of the concert or rehearsal; beginning vs. end. (look for features pertaining to exhaustion) 10. Rehearsal vs. performance 11. Conducting segments vs. non-conducting segments (such as talking or even driving) 12. Noninformative vs. informative gestures while conducting (such as page-turns and stand-adjusting) 13. Surprising or exciting moments in the concert vs. normal moments 14. Posture (verticality of torso) compared with the content of the music; since people tend to lean in when they are attracted and away when they are disgusted, perhaps leaning in or away from the orchestra might have explicit affective content.

A more thorough line of research might use these criteria as a more formal basis for an extended study over many subjects.

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

Acknowledgements

To Professors Tod Machover and Rosalind W. Picard of the MIT Media Lab, whose gracious support, advice, and encouragement have made this doctoral work possible and enjoyable.

To Professors David Wessel of U.C. Berkeley and John Harbison of MIT, whose critical questions have helped me to focus the scope of this project and reexamine my assumptions.

Many thanks go to all those who contributed to the collection of the data! They include: Gregory Harman, Noshirwan Petigara, Jocelyn Scheirer, Eric Scheirer, Young-Moo Kim, Keith Martin, Will Glesnes, Bernd Schoner, Rosalind Picard, Jennifer Healey, James Blanchflower, Susan Anderson, Jim Downey, the Polhemus Corporation, Benjamin Zander, Karen Snitkin, the Youth Philharmonic Orchestra of Boston, Keith Lockhart, Jana Gimenez, Denis Alves, the Boston Pops Orchestra, Stephen Ledbetter, Larry Isaacson, the Greenwood High School Orchestra of Wisconsin, Liling, Augusto, Kelly Corcoran, Dr. Peter Cokkinias and the Boston Conservatory of Music.

References

Auer, V., M. M. Maes, M. C. Bonfim, M. M. Wanderley and M. V. Lamar. 1996. "3D Position Signal Acquisition System with Application in Real-Time Processing." In *Proceedings of the ICSPAT'96 - International Congress on Signal Processing Applications Technology*.

Boie, Bob, Max Mathews and Andy Schloss. 1989. "The Radio Drum as a Synthesizer Controller." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 42-45.

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

Camurri, Antonio, Matteo Ricchetti, Massimiliano Di Stefano and Alessandro Stroscio. 1998. "EyesWeb – toward gesture and affect recognition in dance/music interactive systems." In *Proceedings of XII Colloquium on Musical Informatics*.

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

Clynes, Manfred. 1977. *Sentics: the Touch of the Emotions*. New York: Doubleday and Co.

Darrell, Trevor J., and Alex P. Pentland. "Attention-driven Expression and Gesture Analysis in an Interactive Environment." *MIT Media Laboratory Vision and Modeling Group Technical Report* No. 312.

Harling, P.A., and A.D.N. Edwards. 1997. "Hand Tension as a Gesture Segmentation Cue." In *Proceedings of the Gesture Workshop '96*. London: Springer-Verlag.

Kurtenbach, Gordon, and Eric A. Hulteen. 1990. "Gestures in Human-Computer Communication." In *The Art of Computer-Human Interface Design*. Reading, Brenda Laurel, ed., MA: Addison-Wesley Publishing Company, pp. 309-317.

Lee, M., Guy Garnett and David Wessel. 1992. "An Adaptive Conductor Follower." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 454-455.

Lusted, Hugh S., and R. Benjamin Knapp. 1989. "Music Produced by Human Bioelectric Signals." American Association for the Advancement of Science.

Machover, Tod, and Joseph Chung. 1989. "Hyperinstruments: musically intelligent interactive performance and creativity systems." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 186-190.

Marrin, Teresa, and Rosalind Picard. 1998. "Analysis of Affective Musical Expression with the Conductor's Jacket." Proceedings of the XII Colloquium on Musical Informatics, Gorizia, Italy.

———, and Rosalind Picard. 1998. "The Conductor's Jacket: a Device for Recording Expressive Musical Gestures." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 215-219.

———. 1997. "Possibilities for the Digital Baton as a General-Purpose Gestural Interface." Extended abstract, In *CHI '97 Conference on Human Factors in Computing Systems*, pp. 311-312. NY: ACM Press.

———, and Joseph Paradiso. 1997. "The Digital Baton: a Versatile Performance Instrument." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 313- 316.

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

Morita, Hideyuki, Shuji Hashimoto and Sadamu Ohteru. 1991. "A Computer Music System that Follows a Human Conductor." *Computer Magazine* 24(7), July , pp. 44-53.

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

Picard, Rosalind. 1997. *Affective Computing*. Cambridge, Mass.: MIT Press.

Rovan, Joseph, and Marcelo Wanderley. 1998. "Gestural Controllers: Strategies for Expressive Application." In *Proceedings of SEAMUS '98*.

Rovan, Joseph Butch, Marcelo Wanderley, Shlomo Dubnov and Philippe Depalle. 1997. "Instrumental Gestural Mapping Strategies as Expressivity Determinants in Computer Music Performance", *KANSEI - The Technology of Emotion*, October 1997.

Ryan, Joel. 1992. "Effort and Expression." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 414-416.

Sawada, Hideyuki, Shin'ya Ohkura, and Shuji Hashimoto. 1995. "Gesture Analysis Using 3D Acceleration Sensor for Music Control." In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 257-260.

Usa, Satoshi, and Yasunori Mochida. 1998. "A conducting recognition system on the model of musicians' process." *Journal of the Acoustical Society of Japan,* vol. 19, no. 4, 1998.

Wessel, David. 1991. "Improvisation with Highly Interactive Real-Time Performance Systems". In *Proceedings of the International Computer Music Conference*, San Francisco: International Computer Music Association, pp. 344-347.

Widmer, Gerhard. 1995. "Modeling the Rational Basis of Musical Expression." *Computer Music Journal*, vol. 19, issue 2, Summer 1995, pp. 76-96.