A Compact, High-Speed, Wearable Sensor Network for Biomotion Capture and Interactive Media

Ryan Aylward
Responsive Environments
MIT Media Laboratory
Cambridge, MA USA
1-617-452-5647
aylward@media.mit.edu

Joseph A. Paradiso
Responsive Environments
MIT Media Laboratory
Cambridge, MA USA
1-617-253-8988
joep@media.mit.edu

ABSTRACT

In this paper, we present a wireless sensor platform designed for processing multipoint human motion with low latency and high resolution. One application considered here is interactive dance, in which a choreographer wishes to translate the movements of multiple dancers into real-time audio or video content to accompany the performance. This can only be accomplished using a distributed measurement system capable of responding quickly with enough information to describe the expressive movements of multiple people. Similar requirements exist for biomechanical analysis, especially in the context of athletic training, where high resolution is demanded, and instant feedback is also desirable.

Our approach to addressing such aggressive requirements involves a high-speed wireless network of compact inertial measurement units (IMUs) that can be worn at various locations on the body. Each device is equipped with its own 1Mbps radio link and a full six-axis IMU, as well as a capacitive node-to-node proximity sensor. Currently, the system supports real-time data collection and processing for up to 25 nodes with 100Hz full state updates, thereby handling much higher data rates than its predecessors. With locally buffered data, sample rates of up to 1kHz have been achieved successfully. Early results discussed here demonstrate the feasibility of our design through testing with both dancers and professional athletes.

Categories and Subject Descriptors

B.4.1 [**Input/Output and Data Communications**]: Data Communication Devices. *Transmitters, Receivers, Processors.*

C.3 [**Special-Purpose and Application-Based Systems**]: *Real-time and Embedded Systems.*

H.5.3 [**Information Interfaces and Presentation**]: Group and Organization Interfaces. *Synchronous interaction.*

J.5 [**Arts and Humanities**]: *Performing arts.*

J.3 [**Life and Medical Sciences**]: *Health.*

General Terms

Algorithms, Measurement, Performance, Design, Theory.

Keywords

Wearable sensors, wireless, interactive media, dance, biomechanics, biomotion, real-time, high-performance, inertial measurement unit, synchronous motion analysis.

1. INTRODUCTION

The initial motivation for developing a high performance sensor platform was interactive media. Specifically, our intent was to explore the possibility of generating musical content derived from the collective movement of multiple dancers in real-time. Apart from the merit of artistic innovation, sensing dance poses a significant technical challenge, in that there is a need to capture movement across an unrestricted space, to account for the many degrees of freedom of the human body, to cope with strenuous physical activity without hindering movement, to use sampling rates fast enough to measure rapid motion, and to provide resolution high enough to capture the subtlety of this motion. In addition to these already rigorous constraints, the goal here was to collect, analyze, and transform sensor data into musical parameters in real-time, making continuous transmission at very high data rates a necessity. Therefore, the most critical concern was pushing the limits of available bandwidth, followed by maintaining a compact form factor and reasonable battery life.

Of particular value is the fact that this priority structure differs substantially from that of more common sensor network applications. Typically, for these applications, functionality should be pared down to create the most inexpensive and power-minimal design possible. Much slower phenomena or transient events are measured over longer time periods, data is reduced within the network, and information trickles back slowly as a function of the power limitations and deployment conditions. Since throughput is not often the focus of wireless sensor technology, many recent sensor systems for dance lack the ability to scale to multiple users in a satisfying way. This point will be elaborated below, and is discussed in more detail in [2,3]. Further, any system meeting the challenges of dance is directly applicable to other high performance, immediate feedback applications where the typical sensor network design goals become a limitation. Possible examples include augmenting musical ensembles, adding live media content to team sports broadcasts, professional athletic training, martial

arts, interactive personal fitness monitoring, or physical therapy.

The system described in this paper addresses the goals of improved performance and scalability associated with such applications. Results have been obtained with a dance ensemble, demonstrating simultaneous data collection and fusion in real-time from as many as six dancers wearing four sensors each. In addition, work has been accomplished in the process of extracting features relevant to the collaborative movements of dancers, and mapping these features to musical parameters with low latency. Testing has also been carried out for professional athletic training applications, in which some latency was traded for much higher sample rates and more points of measurement on a single user. The results of this study suggest that high-speed wireless inertial sensors may be used to detect significant information missed by standard video motion capture techniques.

1.1 Sensor Strategies for Dance

Dance is an inherently flexible art form, in that it often involves the addition of other creative elements, including music, video, lighting, or narrative. Naturally, there is a long history in the use of technology to automate or enhance the confluence of these elements [3]. However, as suggested above, challenges still exist in the methods used to sense movement and in the sheer amount of data generated.

As evident by the system discussed here, one general strategy for designing dance interfaces involves placing sensors directly on the body, with a wireless communication link to transmit data to a central computer. Several interfaces of this sort have been developed to capture dance gestures over the last few decades. A number of the earliest of these placed specific emphasis on sensor devices built into shoes, such as the Taptronics system in the 1980's [18], or the Expressive Footwear shoes developed at the MIT Media Laboratory in the 1990's [16]. Unfortunately, shoe-based systems cannot measure upper body or arm motion, and have typically been designed with one dancer in mind. Because of bandwidth constraints, Expressive Footwear was never scaled beyond one pair of shoes, for example.

Attempts to extend wearable dance sensors to other locations on the body have usually started with bendable sensors spanning primary joints, such as the elbows or knees. Architectures of this sort have been introduced at DIEM [19], and by Mark Coniglio of Trokia Ranch [9]. Although systems like these have become wireless, they typically employ a single radio in a central pack, so the various sensors need to be tethered across the body to the wireless unit. This type of design can be very restrictive for a dancer, and the bulky infrastructure tends to limit reliability and scalability in a performance setting.

Another general strategy for sensing dance minimizes or eliminates body-worn hardware, by exploiting computer vision to interpret images received by a camera system monitoring the performance. This technique is now well established, and the prevalence of optical tracking methods has even prompted some artists to develop their own video analysis tools [10,21]. Real-time video analysis is processor intensive, and although the underlying technology and algorithms are steadily improving, computer vision is further limited by constraints on lighting and choreography. Robustness to occlusion and background noise remains problematic, especially with multiple subjects to track. As compared to inertial sensors, sample rates are severely limited in all but the most expensive video systems. Hence, obtaining relevant features reliably from multiple dancers in real-time at the rates desired here is difficult.

1.2 High-Performance Wireless Platforms

Many other compact wireless sensor clusters exist, developed both within our own research group [6] and at other institutes. However, very few, if any systems have combined low power and small size with the number of sensor channels and high data rates needed for real-time multipoint human motion tracking [2,3].

Recently, Flety and collaborators at IRCAM [11] have built wireless sensor networks that use the WiFi 802.11 standard, and have used their WiSeBox system in a dance setting with multiple performers. Although WiFi provides very high data rates, it tends to be much too power hungry for efficient continuous operation with a small battery. Consequently, the WiSeBox architecture is based on a bulky central radio and processing unit, and the sensors themselves have to be wired across the body.

Other new systems favor sensor nodes using Bluetooth for achieving high data rates and ease of connectivity with consumer electronic devices. Some of these use inertial sensors to measure human motion, and have been designed with goals very similar to the ones outlined here [4,13]. Unfortunately, Bluetooth also has relatively high power requirements, and the size of the network is limited to only seven slave nodes per master. This means that Bluetooth places limitations on the scalability of a system for dance. A Bluetooth system with completely encapsulated sensor nodes and high data rates has been proposed in [14], including applications cited for dance and athletic motion analysis. However, to expand this system to multiple users, each person would have to carry a Bluetooth master device with a higher speed uplink. Unfortunately, the equivalent of the heavy central radio pack still exists in this case. The solution proposed in this paper is to dispense with the convenience of Bluetooth or WiFi in favor of a custom protocol using a configurable, high data rate, low-power radio. This allows for fully encapsulated nodes with increased scalability.

1.3 Collaborative Interfaces

One advantage to deploying enough sensing power to track multiple people simultaneously is an opportunity to study collaborative motion, and to design systems capable of responding to relationships between people's activity. This ability is especially desirable, as collaboration is a fundamental basis of human activity. A number of relevant sensor platforms and data analysis strategies geared towards this are discussed in [3]. Despite these initiatives, it seems that only a few attempts have been made to study collaborative performance interfaces from the perspective of detailed collective motion analysis. In the interest of exploring this area more thoroughly, the analysis performed with our system so far has mainly focused on features relevant to group movement.

2. HARDWARE

As suggested above, the system was envisioned as a network of compact sensor devices, each with its own wireless data link, to be worn on the wrists and ankles of a group of users.

For dance, sensor placement on the limbs was chosen as a compromise between the ability to describe full body movements and the number of devices required. Since the nodes are fully encapsulated units, the system can easily be generalized and reconfigured to suit other applications. Each sensor node includes a full IMU, measuring acceleration and rotational velocity in all three axes, with the help of small micro-electromechanical (MEMS) sensors. It was also considered beneficial to include options for extended sensing modalities where resources were available. This led to the inclusion of a capacitive node-to-node proximity sensing system and a flexible expansion port.
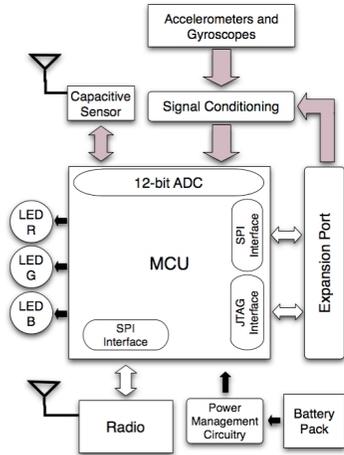

Figure 1. Sensor node architecture.

The basic architecture is shown in Figure 1. Based on the frequency content of active human motion and the peak resolution of the MEMS devices, all channels are sampled at 100Hz and with 12 bits. In order to ensure low latency data collection and synchronous sampling across the network, we use a star topology in which a central base station arbitrates all timing and data collection. The base station forwards the data it receives via USB to a host computer, where feature extraction and analysis takes place. Figure 2 shows the assembled sensor node board, which measures 4x4x2cm. Including the external battery pack, the finished product weighs only 45 grams.
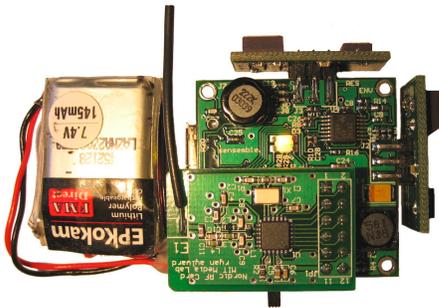

Figure 2. Assembled node board and battery pack.

2.1 Inertial Measurement Unit

The IMU is the functional basis of the system, and hence it is a core component of the hardware design. Accelerometers and gyroscopes of the MEMS variety provide the best precision for compact wireless systems. The choice of specific sensors was heavily influenced by previous work leading up to the current design [7,15].

For maximal range and small size, the Analog Devices iMEMS ADXRS300 was the best rate gyroscope available at the time components were evaluated. However, its nominal range of ±300deg/sec is still low for rapid human motion. Luckily, its sensitivity can be reduced to obtain a range of up to ±1200deg/sec by setting an external resistor [3]. For athletic applications, this measurement range must be extended even further, as explained in Section 4.2.1 below. Although dual-axis gyroscopes have become available recently, the ADXR300 is a single axis device, which necessitates orthogonal mounting of three components to obtain rotational velocity sensing in all three axes. This was accomplished using an edge-mounted daughtercard design, as evident in Figure 2.

Three-axis accelerometers were also new to the market during the early design stages. As sensor daughtercards had already been dictated by the gyroscope choice, we focused on two-axis accelerometers, using an orthogonally mounted pair to obtain the full three axes of acceleration sensing. In order to ensure the full range of measurement for human motion, the ±10g Analog Devices iMEMS ADXL210E was selected.

A number of improvements are already clear with regard to the inertial sensor module. Experimentation has suggested that most dance movement occurs within a ±5g range, allowing for more sensitive accelerometers than the ADXL210E. The new three-axis, ±3g ADXL330 might be adequate, and would not only provide a savings in board area and component count, but would reduce cost and power consumption as well. Adding new multi-axis gyros will be a dramatic improvement, eventually eliminating the need for bulky daughtercards entirely. This progression only highlights the fact that inertial measurement modules are steadily becoming smaller and cheaper.
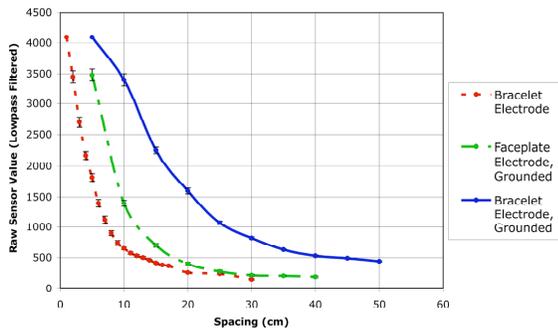
2.2 Capacitive Proximity Sensing

Despite the ability of MEMS accelerometers and gyroscopes to perform well in detailed gesture tracking applications, their signals are still not clean enough to provide positional tracking in space. Because the computation of position from velocity or acceleration requires a double integration, noise and bias rapidly accumulate in the result. To recover from this drift, navigational systems need to be calibrated periodically with an outside reference. In the case of small and relatively noisy MEMS sensors, this recalibration would have to occur on the order of every 5 seconds [7]. Even then, finding a reliable calibration reference is problematic. Since the primary focus here is on relative qualities of movement, energy-based features, and subjective comparisons between gestures, it is possible to leave the issues of calibration and elimination of IMU drift for future work. However, in some cases it would be satisfying to make a simple statement about the location of one sensor node with respect to another. To help extract this type of information, a capacitive sensor system was added to the design, supporting rough measurements of distance between pairs of nodes in the network.

Capacitive sensing comes in a variety of forms, but it usually implies a measurement of capacitance between two electrodes in order to determine electrode spacing [5]. For human interfaces, the electrodes may be irregular conductors worn on the body, or the body itself may act as a ground electrode, and

capacitive coupling occurs through the air. One method of measuring capacitance in this situation is to transmit a high amplitude sinusoidal signal from one electrode, while measuring the received amplitude on the other. For the appropriate transmit frequency range, 50-100kHz in air, the capacitive coupling between electrodes acts as a variable cutoff highpass filter. Therefore, as the capacitance varies, so does the amplitude of the received signal. This configuration is similar to transmit mode sensing as discussed in [20]. Because the variations in capacitance are small and sensitive to environmental factors, synchronous demodulation is typically used to determine the amplitude of the received sinusoid while rejecting uncorrelated noise.

In this case, the transmitter and receiver electrodes are attached to isolated wireless sensor nodes. Because of this, there is no direct electrical path between the two devices, which makes the situation more challenging. To provide a common reference, the devices worn by the same user can be grounded through the body, a technique that has been documented to improve performance in [17]. Another difficulty is the lack of phase coherence between the receiver and transmitter, which prevents synchronous demodulation from being performed. However, the amplitude of the received signal can still be estimated by sampling in quadrature, or at exactly four times the transmit frequency, as established in [12] and further discussed in [3,20].

Our design uses an amplifier and tuned LC oscillator driven by a microcontroller output pin to generate the transmit signal at 90.1kHz. Peak-to-peak amplitudes of up to 40V can be achieved with this technique, with very little power drain. When receiving, the node is able to use the same LC oscillator as a bandpass amplifier. Since only one capacitive measurement can be made at a time, nodes are designed to continually trade roles as transmitters and receivers. A detailed discussion of the implementation is provided in [3].

Figure 3. Typical capacitive sensor response with different electrode configurations.

This capacitive sensor design was first presented in [1], but since then there have been significant improvements in range and reliability. Figure 3 shows the performance of the system with different electrode configurations, and with the body grounded and ungrounded. The faceplate electrode indicated in the chart was a small piece of copper foil spanning the area of the node circuit board. The bracelet electrode was a wide copper band encircling the wrist, used to simulate the effect of building a conductive electrode into the strap affixing the sensor device to the body. As expected, provided that two nodes share a ground through the body, the bracelet electrode provides a significantly better range than the smaller faceplate electrode, since capacitance is dependent on electrode area.

However, eliminating the shared ground cuts this range nearly in half. In fact, the results indicate that a small electrode employing a shared ground is preferable to a large electrode with no shared ground. Assuming a favorable electrode design, the current capacitive sensor system can achieve a range of about 50cm. This covers gestures occurring across most of the range of a typical arm span, and the symmetrical bracelet design is able to eliminate the directionality issues of infrared or ultrasonic ranging systems.

2.3 Radio

The wireless radio on each sensor node is the most critical component of the hardware design next to the inertial sensor module. Our goal was to achieve the highest data rate possible with a low-power radio, in order to enable higher data throughput than other wireless tracking systems. To establish the bare minimum requirements, consider five dancers wearing 4 sensor nodes each. With 100Hz sampling rates and 12 bit samples as stipulated above, this produces 144kbps of inertial data alone:

6sensors/node $\times$ 12bits/sensor $\times$ 20nodes $\times$ 100Hz = 144kbps.

Given a rough estimate of 40% channel utilization due to overhead and protocol inefficiency, as well as the addition of four additional sensor signals possibly generated by the capacitive sensing system, the actual bulk data rate requirements of the communications channel can be much higher:

10sens/node $\times 12$bits/sens $\times 20$nodes $\times 100$Hz $/40\%$ = 600kbps.

During the early design stages, the Nordic nRF2401A was the fastest option. This is a low-power 2.4GHz radio providing data rates of up to 1Mbps. The radio module was designed on a small daughtercard, physically separate from the main node board, and equipped with an SPI interface for communication with the microcontroller. A daughtercard module was favored for signal integrity and interchangeability. The sensitive nature of RF circuitry is such that it is preferable to isolate the radio from the rest of the system to minimize interference. As far as interchangeability is concerned, the nRF2401A has already been superseded by the potentially more flexible 1Mbps Chipcon CC2400 and the new 2Mbps Nordic nRF24L01 during the course of the development of this project. Judging by this trend, capabilities of low-power radios will continue to improve rapidly for the foreseeable future, mirroring the improvement of inertial sensing devices. By encapsulating the radio module on a replaceable daughtercard, the sensor node can adapt easily to higher network speeds as possibilities grow.

2.4 Microcontroller

Each sensor node is equipped with a microcontroller (MCU) which functions as a local control center for collecting and processing data, arbitrating sensor behavior, maintaining communication with the RF module, and timing events. Although there are data radios with integrated MCUs, including the 1Mbps Nordic nRF24E1, we favored a dedicated MCU for flexibility, processing power, and increased I/O. In the current design, the Texas Instruments MSP430F149 was used. This is a low-power, 8 MIPS, 16-bit device with an 8-channel, 12-bit ADC, ideal for small sensor systems. The MCU provides some spare resources, including a free digital input for interfacing with a Polar heart rate monitor, a free SPI

interface for connecting with other digital devices, and a free analog input with associated signal conditioning circuitry for handling an additional resistive sensor, such as a pressure sensor, bend sensor, or light sensor. All of these signal lines are broken out to the expansion port shown in Figure 1, which also acts as the programming interface.

2.5 Power

Power is supplied by a 7.4V 145mAh lithium polymer battery pack. This chemistry was chosen because it is lightweight, compact, and rechargeable. Unfortunately, continuous operation with three gyroscopes and frequent radio usage prevents the design from meeting traditional low-power requirements. During normal operation, each node can consume over 200mW. Therefore, the lithium battery pack allows for about 5 hours of operation, which was considered adequate for a dance performance or motion capture session. The battery pack was decoupled from the sensor board in order to distribute the mass of the device, thereby making it more comfortable to wear. Also, access to the battery makes it easy to swap out for charging and allows for the use of other battery configurations in the future.

3. COMMUNICATIONS

Data generated by the sensor nodes must be transmitted to the central computer as quickly as possible to allow real-time processing to occur. In the case of music generation, delays of greater than 100ms are not only clearly audible, but can be disruptive to performers. Because of this, transmission latencies of greater than 30ms begin to leave insufficient time for processing. Additionally, the samples across all of the sensors in the network should be time synchronized and collected at a stable rate. As suggested previously, the best arrangement in this situation is a star topology, where a base station is given complete control over all communication and timing on the network.

In the custom protocol proposed here, which has been largely derived from an earlier design [15], this control is based on broadcast messages sent from the base station to all of the nodes at the sample rate of 100Hz. This imposes very strict synchronization, and allows the nodes to share the communications channel effectively with a simple time division multiple access (TDMA) scheme. At the beginning of the cycle, each node listens for the broadcast signal. Upon receiving the signal, the sensors are sampled, and data is transmitted back to the base station after a preprogrammed time interval, as determined by each node's hard-coded ID number. Once a transmission has been sent, each node knows roughly when to begin listening again in preparation for the next broadcast (see Figure 4). During the time when the radio is not transmitting or receiving, it is put to sleep to conserve power. At the base station, the data received on each cycle is forwarded to a host computer via a USB connection. Besides low latency, the advantage of this protocol is that samples across the entire network will be taken at roughly the same time and precisely at the sample rate determined by the base station, with high tolerance to clock skew between nodes.

Data is sent upstream from each node to the base station in 16 byte packets for each sample, which include a header, 6 inertial sensor values at 12 bits each, and 3 capacitive sensor values at 16 bits each. During transmission, the radio also adds an 8 bit preamble, 40 bit address, and 16 bit cyclic redundancy check (CRC) code, for a total packet length of 24 bytes or 192 bits.

The transmission also requires a setup time of roughly 195μs in addition to the 192μs air-time required to send 192 bits at the rate of 1Mbps. Therefore, the effective data throughput for a single transmission with this configuration is only 330kbps, or a 33% channel utilization:

16 bytes / (195μs + 192μs) = 128bits/387μs = 330kbps.

In practice, the setup times for transmission on different nodes can be overlapped to pack received messages at the base station as tightly as possible. This helps recover some loss, but other sources of inefficiency are present, including the time within each broadcast cycle dedicated to transmitting a broadcast message, collecting new samples on the nodes, and processing the received samples on the base station. The system also lacks efficiency due to the fact that it must handle many small packets, each with overhead, rather than a smaller number of large packets. Based on empirical measurement, a maximum of 25 nodes are able to share the channel simultaneously while sampling 100Hz, given the current protocol and packet configuration. Hence, the actual data throughput of the sensor network is found to be:

25nodes x 128 bits/node x 100Hz = 320kbps.

Although the loss in data rate appears drastic compared to the bulk rate of 1Mbps, this efficiency is representative of what could be achieved with standard protocols, and our custom scheme has the advantage of low latency, with a maximum of one sample delay to the base station and typically a two sample delay to the host computer. More details on the communications structure can be found in [3].
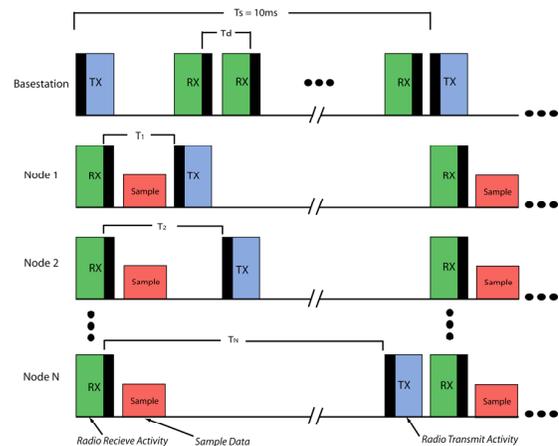

Figure 4. Illustration of the TDMA communications cycle.

The success of this system in the field depends on the RF environment. In general, signals were received very reliably over a range of about 50 feet indoors. However, the existence of nearby 2.4GHz activity, as well as shielding from the body, can disrupt communication within this range, causing received packet rates to drop to 70% in some cases. One problem is that the protocol is highly dependent on broadcast packets being received by every node at every sample interval. In fact, clock skew between nodes is small enough that the network can continue to return data within the TDMA slots for at least two samples per resynchronization. Making this simple optimization in the future will greatly reduce packet loss in cluttered RF environments. In this case, dance performance spaces tend to limit the problem of interference, as the environment on stage can be somewhat controlled.

4. ANALYSIS AND RESULTS

4.1 Feature Extraction

Early experimentation mainly focused on establishing a set of low-latency features that could be generated from our sensor data with meaningful interpretations in a dance context. In particular, we wished to highlight the unique ability of our system to analyze relationships between the movements of multiple people. This research was initially presented in [1,2]. The two basic concepts explored in terms of collective motion analysis are measuring time separation and spatial similarities between movements and monitoring and comparing activity levels. In addition, the possibility of exploiting the structure of a group to mediate feature extraction has been introduced.

4.1.1 Temporal and Spatial Separation

In the context of a dance ensemble, time and spatial correlations can be used to determine which dancers are moving together, which dancers are leading versus following, and perhaps which are responding to one another with complementary movements. The frequency of repetitive motion can also potentially be detected, leading to an estimate of mean tempo. Similarly, in situations outside of dance, the ability to document how the movement of one person affects that of another can be valuable in determining the nature of the collaboration taking place.

Cross-correlation is the most direct way of obtaining this information, as the location of a cross-correlation peak denotes the time delay between similar events, and the height of the peak denotes the degree of similarity. However, since the accelerometer signals tend to have a bias offset due to gravity, cross-covariance is the preferred measurement, as it is the equivalent of cross-correlation with the mean of the input signals removed. On streaming data, cross-covariance calculations can be performed periodically on short windows of data. Window size is chosen to make a tradeoff between latency and the maximum time separation that can be expressed. For instance, a window of length N samples handles time separations of $\pm N$, but also requires N samples before the calculation begins. Therefore, cross-covariance can never be extracted in real-time, but processing on windows less than a second long might still be useful for driving interactive content.

To test the effectiveness of a cross-covariance measure, sensors were given to three dancers participating in a ballet lesson; each wore one on the right ankle. The class then performed an exercise involving a repeated sequence of leg swings executed in unison, to music. Although they were roughly in time with the music, the dancers were not necessarily looking at each other or at an instructor, creating a small but clearly visible delay in their motions, as documented on video.

Figure 5 shows the result of windowed cross-covariance analysis on the corresponding data segment, computed using a window size of 1 second, step size of 0.25 seconds, and averaged across sensor axes. That is to say, at each interval of 0.25 seconds, data from the past second was considered, the cross-covariance vector was computed individually for each inertial sensor signal, and finally the individual vectors were averaged to produce the final result. The segment from about 35 to 65 seconds corresponds to the synchronized sequence of swings made with the right leg. Note that the area of peak cross-covariance, shown in white, tends to waver around the

baseline as time progresses. This is consistent with the dancers slowly leading and lagging with respect to one another by small amounts. The histograms in Figure 6 roughly show the extent of the peak drift for each of the three plots in Figure 5. It is clear from the relative stability of middle plot that Dancer A and Dancer C were closely synchronized for the duration of the exercise. However, the other two pairings were not as stable. These fluctuations reflect accurately what is visible in the video. Interestingly, it turns out that Dancers A and C were facing each other during the exercise, while Dancer B had her back turned to the others.
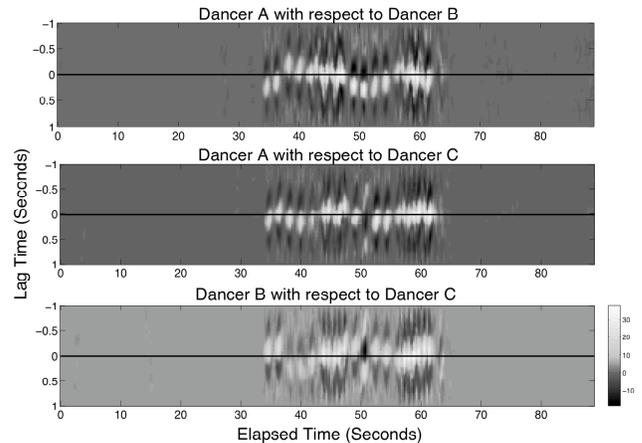

Figure 5. Average windowed cross-covariance between pairs of dancers, for leg swings performed in unison.
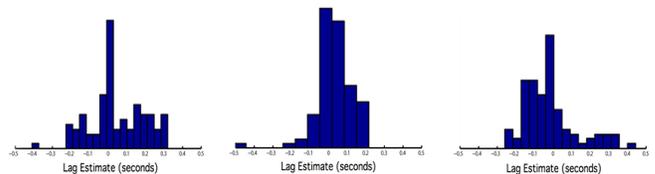

Figure 6. Histograms showing XCOV peak drift in each of the three plots of Figure 5, top to bottom.

Another observation regarding Figure 5 is the fact that in some areas, the covariance peaks are not as well defined as others. This is especially true in the bottom plot, relating dancers B and C. Smaller peak values indicate less similarity in the motions being performed, and this data seems to agree with certain gestural differences visible in the video footage. Thus, although cross-covariance is not strictly a real-time feature, it is still clearly valuable for describing group relationships, both temporally and gesturally.
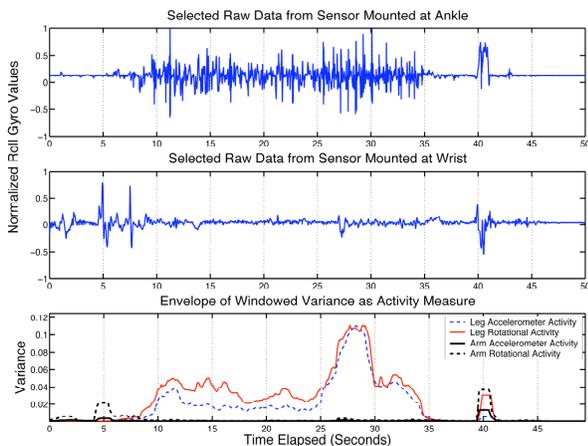
4.1.2 Quantifying Activity

In addition to extracting correlations between the activities of a group, it is important to obtain information about the properties of the activities being observed. These properties might include variations in the overall activity level of an individual or group at different time scales, principal axes of movement, or other features extracted during an interval of high activity. Activity related measures can also be calculated with lower latency than cross-covariance.

Increased physical activity, as qualified by faster movements and more frequent directional shifts, is related to the energy present in the inertial sensor signals. In turn, the average energy over a segment of data is reflected in the variance of the

segment. Therefore, a common approach to activity measurement on an inertial data stream involves computing windowed variance. The associated computations require very little processing power, and although they also rely on capturing a complete window of data, these segments can be much shorter than what was needed for cross-covariance. Hence, the latency associated with windowed variance is much lower. Variance can be used here with various combinations of sensors, and can be processed in different ways, depending on the desired result. For instance, if the separation between gestures is long enough, the variance spikes created at the beginning and end of a movement can be used to delineate them. In other cases, it might be useful to use a median filter to obtain a slowly varying envelope on the running variance for certain sensors, in order to determine broader trends in activity level.

As an example of the latter, data was collected from the right wrist and ankle of a ballet student performing a sequence of motions in which slow kicks with the right foot transitioned into fast, tense kicks. The full sequence was framed with a stylistic raising and lowering of the right arm at the beginning and end, respectively. Figure 7 shows a portion of the raw data from this segment along with four different activity envelopes obtained by filtering the windowed variance of both upper and lower body movement. Accelerometer activity here denotes the average variance envelope across the accelerometer axes, while rotational activity denotes the average across the gyro axes. One can clearly see a marked increase in activity as leg motion transitions to faster kicking. The role of the arm movement is apparent in the activity envelope as well.

Figure 7. Selected data and resulting activity envelopes as dancer transitions from slow kicks to rapid, tense kicks. Sequence of leg motions is framed by stylistic arm motion.

If one wishes to identify a particular type of activity, it may be more important to compare motion along each sensor axis than comparing rotational versus translational motion, since a specific movement may be characterized by high variance in some directions, but not in others. Previous work in [2,3] discusses how variance envelopes from multiple sensor axes on one individual could be combined to form an activity profile for distinguishing between certain types of movement. This type of analysis can be extended to an entire ensemble, simply by averaging the appropriate activity envelopes across the group to create a global activity feature. Global activity can be useful for determining predominate axes of collective motion across the group.

4.1.3 Feature Reduction from Group Structure
In order to limit the amount of information that has to be processed and interpreted to generate meaningful feedback, it is helpful to reduce the number of high-level features under consideration at any given time. In any situation with group cooperation, movement develops a structure that may create very natural opportunities to make this reduction.

In the case of dance, one possibility is the use of lower level features such as activity profiles to cluster groups of dancers. Ideally, these clusters would reflect real groupings forming on stage as a result of the choreography. Once the clusters have been made, heavier analytical techniques can be performed on these groups as units, rather than on each dancer individually.

Another option is to consider data selectively based on the statistical properties of the group. This might involve giving special analytical attention to individuals whose movements deviate significantly from the norm, or it might involve focusing only the net characteristics of the ensemble, giving weight to the majority. These structural interpretations can help narrow down the space for mapping and interpretation.

4.2 Biomotion Capture for Athletics
Although interactive dance provided the impetus for this sensor system, its possible uses extend to any area in which high-quality human motion analysis can be applied, especially those requiring low latency feedback, high sample rates, or many points of measurement. Athletic training comes to mind as one of the most promising areas for further development. In a pilot study detailed in [8], we were able to test our system with a minor league baseball pitcher, in collaboration with the Massachusetts General Hospital Sports Medicine Department and the Boston Red Sox. The motivation for focusing on pitchers stems from a recent increase in injuries associated with the shoulder and elbow of the pitching arm. These joint injuries are typically the result of wear building up over time, but the risk of injury clearly escalates if a pitch is made with poor body mechanics or after passing the threshold of muscle fatigue [8]. A definitive risk assessment is difficult to make, in part because the mechanics of the arm during a baseball pitch are not fully understood. Especially for professional players, who routinely throw fastballs with release speeds approaching 100mph, the critical portion of the arm motion occurring immediately before ball release is simply too fast to be captured in detail with the standard video tracking techniques. In particular, health practitioners are looking for more accurate measurements of peak angular velocities at the shoulder and elbow, as well as peak accelerations at the wrist.

4.2.1 Adapting the Design
Based on rough estimates available in the biomechanics literature, internal rotation of the shoulder was expected to peak at values of up to 10,000deg/sec, elbow extension up to 4,500deg/sec, and wrist acceleration above 80g [8]. In order to measure these extreme ranges, modification of the IMU was required. Because of varying requirements at different locations on the body, high-range and mid-range IMUs were designed to accompany the original low-range design. The mid range device incorporated three ADXL78 ±70g single-axis accelerometers, while the high-range device used a similar arrangement of ADXL193 accelerometers, providing measurement to ±120g. Both modified IMUs were intended to measure rotational velocities up to 10,000deg/sec, which is

well above the nominal rating of any MEMS gyroscope. Luckily, adapting the ADXRS300 to such an enormous range is only a matter of reducing its sensitivity, by overriding some of its internal power management, according to advice from Analog Devices [3]. Using this technique, we were able to increase the range to over 11,000deg/sec.

In addition to high range measurement, high sampling rates were also required to capture motion in sufficient detail. During a baseball pitch, the period of high acceleration over which extreme measurements are registered typically only lasts 20-40ms. Using the standard 100Hz sampling protocol, this equates to only 2-4 data points. State of the art video motion capture systems can double this number of data points; for instance, the cameras used by the American Sports Medicine Institute sample at 240 frames/sec. However, inertial sensors are capable of responding at much higher rates. In order to highlight this ability, we adapted our system to sample at 1kHz, thereby generating data with 4 times more temporal resolution than leading video motion capture systems. This improvement is made at the expense of real-time operation, as the radio does not have the bandwidth required to stream data at this rate from multiple nodes. Instead, data is logged locally in on-chip flash memory and is returned to the base station when a pitch is complete.

4.2.2 Preliminary Results

The study was conducted with a pitcher wearing six sensor nodes placed at the most crucial locations on the body. Table 1 describes the sensor placement and distribution of measurement capabilities. At the upper arm and wrist, pairs of low and high range sensors were used to provide additional resolution for fine movements at low ranges. A video motion capture system was also being tested concurrently with our wireless sensors. Although there has not been a formal attempt to compare the output of the two systems, we hope to validate our system against motion capture data in future work. A limited number of measurements were taken in this study, but compelling results were provided by the high-range accelerometers.
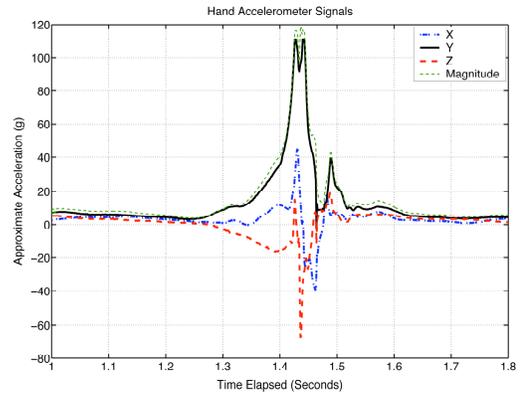
Sensor Position	Accelerometer Range (g)	Gyro Range (deg/sec)
Torso	±10	±1,200
Arm (Low Range)	±10	±1,200
Arm (High Range)	±70	±11,700
Wrist (Low Range)	±10	±1,200
Wrist (High Range)	±120	±11,700
Hand	±120	±11,700

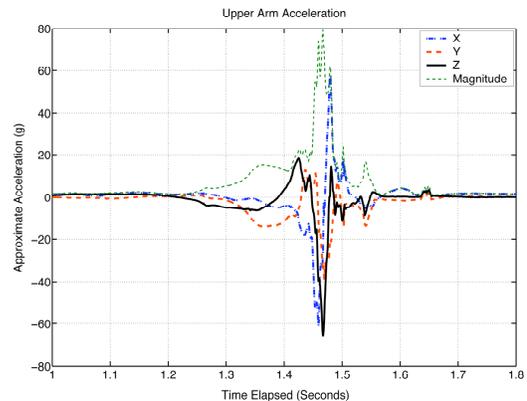**Table 1. Sensor placements and approximate ranges.**

Figures 8 and 9 show an example of the accelerometer signals from each axis as well as estimated magnitudes recorded from the hand and upper arm during one of the pitches. The results indicate an area of significant movement lasting roughly 400ms, with a rapid peak acceleration phase lasting under 30ms, as expected. The acceleration values given in the plots are approximate, as the sensors were never calibrated. Nevertheless, a very sharp increase in magnitude of acceleration was recorded on all nodes, with approximated peak values in accordance with the video tracking results and predictions in the literature.

At the hand, the predominant axis of acceleration is the y-axis as shown by the solid line, which peaks above 110g (Figure 8). This represents the acceleration directed down the arm towards the elbow, and hence is the major component of centripetal

acceleration as the arm rotates towards ball release. The smaller accelerations on the other two axes are transverse movements, as indicated by an acceleration (positive) peak followed by a deceleration (negative) peak. This most likely corresponds to the sudden burst in tangential velocity immediately before ball release. The magnitude plot indicates that at its peak, the hand may experience up to 120g of net acceleration, although this impressive peak is only sustained for several milliseconds. Data recorded during other trials exhibits a similar behavior and peak acceleration value, although more extensive testing is called for before definitive claims can be made.

Figure 8. Acceleration phase of a baseball pitch as measured at the hand.

Figure 9. Acceleration phase of a baseball pitch as measured at the upper arm.

In addition to peak accelerations, it was also important to try to obtain a preliminary estimate as to the peak angular velocity of internal rotation of the shoulder. This process generates most of the power associated with a baseball pitch. Consequently, the shoulder is the joint that takes the most wear from repeated pitching. Unfortunately, the gyroscopes that would have measured this rotational speed directly malfunctioned during the study. Still, it was possible to obtain a rough estimate by assuming that angular velocity of the shoulder is related to centripetal accelerations of the arm. At the upper arm, internal rotation of the shoulder shows up as centripetal acceleration directed inwards towards the bone. This corresponds to a negative acceleration on the z-axis of the upper arm sensor, which is clearly apparent in Figure 9. The centripetal acceleration appears to reach approximately 66g.

Given a biceps radius of 6cm, this results in a peak angular velocity of 104 rad/sec, or roughly 6,000 deg/sec. This is lower than the expected peak value of 10,000 deg/sec, but it is a very vague estimate that may have been affected by the lack of calibration of the accelerometers.

Although the accuracy of our observations is unknown, they seem to agree with predictions of extremely high accelerations and angular velocities that peak for only tens of milliseconds. The data recorded using our system shows that these peaks have definite structure, which cannot be captured reliably at current video tracking rates. With calibration and additional testing, a wireless inertial sensor system could be an important supplement for video motion capture in a sports biomotion context. Once the relationship between pitching parameters and inertial features is more thoroughly studied, it may be possible for an inertial system to provide useful information without optically tracked reference points, providing the potential for bringing biomechanical analysis out of the laboratory and onto to the playing field.

4.3 Generating Interactive Content

As of yet, the sensors have not been used in a live interactive media setting where dancers can actually hear the audio output they create. However, all of the necessary components are in place, and the full system has been tested successfully in an environment that allows sensor data to be realized as sound in a simulation of real-time operation. To accomplish this, five dancers were monitored, each wearing sensors on both wrists and ankles, for total of 20 sensor nodes. Data was collected during a rehearsal over several repetitions of a short dance piece. Then, the recorded data was analyzed offline to determine relevant features and mapping strategies for turning these features into interactive sound. Finally, the recorded performances could be recreated by playing the data back into the feature extraction engine, just as if it had arrived in real-time. All data processing, feature extraction, and mapping algorithms were run in Max/MSP, with sound synthesis in Reason.

A critical part of this test application was to design a low-latency mapping to musical parameters. The goal was not to create a coherent performance piece, but to verify the capabilities of the system, to show a clearly audible relationship between movement and responsive sounds, and to demonstrate how the features discussed previously could be interpreted in a real dance setting. The mapping process is as shown in Figure 10. The feature extraction algorithms employed are meant to cover most of the strategies for group analysis, including temporal separation, comparison of activity profiles, and average group parameters. For simplicity, only the motion data provided by the inertial sensors was considered. During the mapping process, these time varying features are transformed into control parameters for sound generation using four basic synthesized instruments. For this design, the mapping from features to sound parameters is very simple and direct, so that the features being measured are audible.

The most prominent components of the output were guitar and flute sounds triggered by "solo" movement, defined as activity for one dancer that was not reflected in the rest of the ensemble, and violin notes that responded to increasingly synchronized and correlated activity across the ensemble. In the case of the violin notes, as the degree of correlation increases across the ensemble, the density of the triggered

notes increases as well. At the same time, the degree of synchronicity across the ensemble, as measured by average lag times between similar movements, controls the pitch of the notes. These parameters combine to create very frenetic, high-pitched clouds of violin notes whenever the ensemble locks into simultaneous synchronized movement.
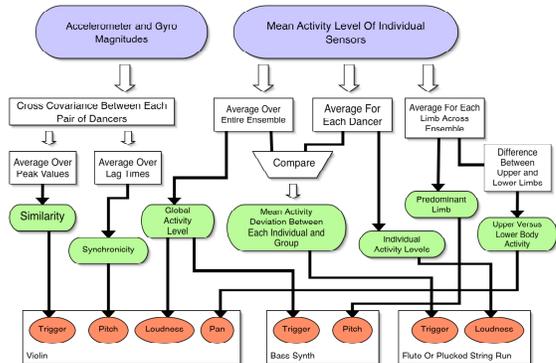

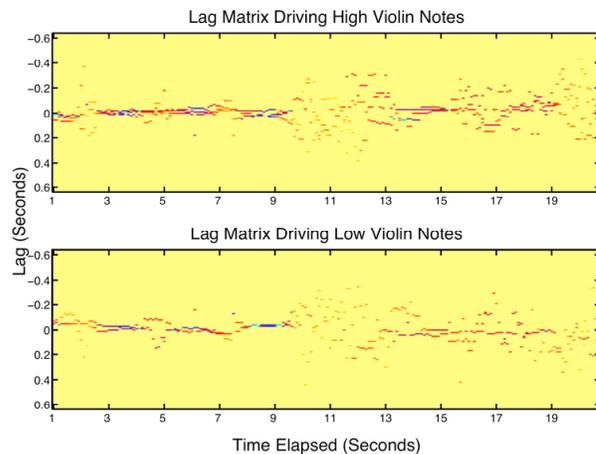**Figure 10. Basic strategy to test real-time mapping from sensor data to sound.**

Figure 11. Average lag features between dancers, used to determine violin pitch and note triggers.

Figure 11 shows the progression of the lag feature used to control violin notes during a particularly clear segment of dance. Dark points correspond to higher pitches, while dispersion from the center axis roughly corresponds to note density. The two plots are indicative of the fact that half of the dancers controlled a violin with a high note range while the other half controlled a lower note range, to prevent clutter. In this segment, the dancers alternate between swinging their arms energetically and running in place. During the arm motion, all of the dancers are very tightly synchronized. However, when running in place the footsteps were more difficult to synchronize, hence there is dispersion along the lag axis. A similar dispersion is heard in the cloud of violin notes at these instances.

Qualitatively, much of the audio content rendered in this manner matched up with dance movements in the recorded video segments in a satisfying way. Some noticeable latency was inherent in the features being used, for example, over a second delay for cross-covariance. However, the mechanics of

the real-time simulation held up well, in the sense that data arriving at realistic rates was translated from features into sound successfully on one processor. Latency will surely be worse in a live situation, and it remains to be seen how dancers respond and adapt to this. With interactive feedback, the ability to utilize low-latency features while still achieving a satisfying output will improve.

5. CONCLUSIONS AND FUTURE WORK

In summary, this paper has presented the design of a system of compact, wearable, wireless inertial sensing devices, as well as their application in analyzing human motion and providing real-time feedback. With dance performance applications in mind, the design effort has been focused on building a platform for improved scalability, speed, distributed measurement, and interpretation of group interactions. The novel achievement in this area is the ability to instrument entire ensembles without sacrificing the measurement resolution typically reserved for an individual. With this requirement satisfied, the result is a high-performance system equally applicable to any situation where human motion analysis is coupled with the need for high data rates.

The results of the dance study show that it is possible to collect and analyze data quickly enough to generate meaningful feedback in response to dance with tolerable latencies. At this stage, the data has mainly been interpreted at a low level. More work is required to find features and processing algorithms that reduce latency while improving the meaningful quality of the output. The next step in this process is to run the system in a live situation where dancers can hear the musical feedback they generate. While playback simulation was arranged to project true latency estimates, it does not replace the ability of dancers to respond, learn, and offer suggestions for improvement.

Extending beyond dance applications, the system has also been evaluated in a preliminary study measuring the arm movement of professional baseball pitchers. With some modifications, the sensors were able to measure the extremely high accelerations involved with more temporal resolution than a state of the art motion capture system. Statements as to the accuracy of these measurements and their bearing on traditional motion capture practices will require further study and rigorous calibration.

As technology continues to improve, wireless inertial measurement modules will certainly become smaller, faster, cheaper, and more accurate. Thus, the strategies developed with this platform will become more feasible and ultimately more prevalent in high-performance motion tracking applications.

6. ACKNOWLEDGMENTS

7. REFERENCES

[1] Aylward, R., Lovell S.D., and Paradiso, J.A. A compact, wireless, wearable sensor network for interactive dance ensembles. *Proc. of BSN 2006*, pp. 65–70, April 2006.

[2] Aylward, R. and Paradiso, J.A. Sensemble: A wireless, compact, multi-user sensor system for interactive dance. *Proc. of NIME 06*, pp. 134–139, Paris, France, June 2006.

[3] Aylward, R. *Sensemble: A Wireless Inertial Sensor System for Interactive Dance and Collective Motion Analysis.* M.S. Thesis, MIT Media Laboratory, Cambridge MA, 2006.

[4] Barry, M. et al. Multimedial enhancement of a butoh dance performance - mapping motion to emotion with a wearable computer system. *Proc. of MoMM*, 2004.

[5] Baxter, L.K. *Capacitive Sensors: Design and Applications*. John Wiley and Sons, Ltd., Canada, 1996.

[6] Benbasat, A.Y. and Paradiso, J.A. A compact modular wireless sensor platform. *Proc. of IPSN 2005*, pp. 410–415, Los Angeles, CA, April 2005.

[7] Benbasat, A.Y. *An inertial measurement unit for user interfaces*. Master's thesis, MIT Media Laboratory, September 2000.

[8] Berkson, E. et al. IMU arrays: The biomechanics of baseball pitching. To appear in *The Orthopaedic Journal at Harvard Medical School*, October 2006.

[9] Coniglio, M. *The MidiDancer system*. http://www.troikaranch.org/mididancer.html.

[10] Downie, M. *Choreographic the Extended Agent: Performance Graphics for Dance Theater*. PhD thesis, MIT Media Laboratory, September 2005.

[11] Flety, E. The WiSe box: A multi-performer wireless sensor interface using WiFi and OSC. *Proc. of NIME 05*, pp. 266–267, Vancouver, Canada, May 2005.

[12] Grace, O.D. and Pitt, S.P. Sampling and interpolation of bandlimited signals by quadrature methods. *J of the Acoustical Society of America*, 48(6A):1311–1318, December 1970.

[13] Kunze, K. et al. Towards recognizing tai chi - an initial experiment using wearable sensors. *Proc. of IFAWC*, 2006.

[14] Laurila, K. et al. Wireless motion bands. *In Ubicomp 2005 Workshops: Ubiquitous computing to support monitoring, measuring, and motivating exercise*, 2005. http://seattleweb.intelresearch.net/projects/ubifit/papers/w10-p3.pdf.

[15] Lovell, S.D. *A system for real-time gesture recognition and classification of coordinated motion.* Master's thesis, MIT EECS, January 2005.

[16] Paradiso, J.A. et al. Design and implementation of expressive footwear. *IBM Systems Journal*, 39:511–529, October 2000.

[17] Paradiso, J.A. and N. Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, 21(2):69–89, 1997.

[18] di Perna, A. Tapping into MIDI. *Keyboard Magazine*, pg. 27, July 1988.

[19] Siegel, W. and Jacobsen, J. The challenges of interactive dance: An overview and case study. *Computer Music Journal*, 22(4):29–43, 1998.

[20] Smith, J. et al. Electric field sensing for graphical interfaces. *IEEE Computer Graphics and Applications*, 18(3):54–60, 1998.

[21] Wechsler, R. et al. EyeCon - a motion sensing tool for creating interactive dance, music, and video projections. *Proc. of the SSAISB Convention*, Leeds, England, 2004.