

A Compact, Wireless, Wearable Sensor Network for Interactive Dance Ensembles

Ryan Aylward

S. Daniel Lovell

Joseph A. Paradiso

*Responsive Environments Group
MIT Media Laboratory
20 Ames St.
Cambridge, MA 02139
joep@media.mit.edu*

Abstract

We describe the design of a compact, wireless sensor module meant to capture expressive gestures in real-time when worn at the hands and feet of a dancer. Each sensor node includes a 6-axis inertial measurement unit (IMU) that contains three orthogonal gyroscopes and accelerometers in order to capture local dynamics, as well as a capacitive sensor to measure node-to-node proximity. The nodes are designed to be easily augmented with other digital or analog sensors. The onboard radio transceiver supports a power-efficient, high-speed RF network capable of real-time data acquisition from several devices simultaneously, thereby meeting the requirements of instrumenting a small dance ensemble for interactive applications. This paper describes the application goals, presents the prototype hardware design, introduces concepts for feature extraction, and shows early test results.

1. Introduction

Many wireless interfaces have been developed to capture dance gesture over the last decade or two. Several have been built into shoes, such as the 1980's Taptronics, featuring piezoelectric pickups at the toe and heel [1] and the Expressive Footwear by our group at the MIT Media Lab [2]. Originally realized in 1997, this system was an early implementation of a dense, multimodal wireless sensor cluster (now becoming common in sensor networks) that measured 16 variables including many degrees of both contact and free-gesture control. Other examples of wearable dance instrumentation typically use bendable sensors that span primary joints, e.g., the elbows and knees. Architectures of this sort have been introduced by DIEM in Copenhagen [3] and by Mark Coniglio of Trokia Ranch in New York [4]. Although these systems have become wireless, they employ a single radio in a backpack or backpack, hence the various sensors need to be tethered across the body to this central dispatcher. Extreme (and expensive) versions of these types of wearable joint-bend interfaces can be found in full-body motion capture outfits (Goniometers) for computer graphics, such as exoskeletons and flexible fiber-optic angle-sensing systems such as the ShapeWrap by Measurand [5].

The systems above were developed for solo performers, and many do not scale well to ensemble performances. For instance, the Expressive Footwear system used all bandwidth provided by the on-shoe radios at their given transmit frequencies to achieve 60 Hz full-state updates. As its basestation sported a pair of receivers, it was only capable of addressing two shoes (although several performances were done with a pair of dancers wearing one active shoe each), and no provision was included to sense upper body and arm motion. Some of the centralized backpack systems enable more than one dancer to be accommodated, but again, wires need to be run from the various sensor locations to the central body-worn data concentrator or transmitter.

Another approach to gesture tracking for dancers avoids any body-worn hardware and exploits computer vision, processing video from a camera (or cameras) watching the stage. This technique has been established for over a decade now, and systems like the Very Nervous System [6], Eyes Web [7], Big Eye, and Jitter are used by many composers – others, (e.g., [9]) have developed their own video analysis tools. This approach 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 clutter and dynamic background remains problematic, and obtaining multiple relevant features reliably from a dance ensemble in real-time can be difficult.

Accordingly, we have developed a compact and relatively unobtrusive wireless sensor package with the aim of instrumenting the hands and feet of a small dance ensemble for real-time collective gesture tracking and musical feedback. This paper details our approach, presents results from current hardware, and describes work in progress on feature extraction for describing ensemble activity.

2. Goals

The motivation for this project is the fairly recent opportunity to leverage low-power, high-bandwidth RF solutions and compact inertial sensors to create a wearable wireless motion sensing system meeting the demands of many points of measurement and high data rates. Our goal is to implement such a system for an

interactive dance ensemble, which is in some ways an ideal situation for pushing high performance requirements. A highly active environment of human motion demands an unencumbering yet sturdy wearable design. Obtaining detailed information about the movement of the human body and the interaction of multiple human bodies demands many points of measurement. Most importantly, using this information as a vehicle for interactive performance, specifically with musical feedback, demands real-time data collection and reduction (e.g., order of 100 Hz full state updates). In the broader scope, we hope to test the applicability of this system to other applications, such as analyzing the dynamics of team sports, physical therapy, biomotion measurement & analysis, and personal physical training.

3. Hardware Design

The main successor to the Expressive Footwear project up to now has been the Sensor Stack [9]. This is a modular wireless sensing cluster developed by our group that can be configured with a variety of small stacking sensing panes, each of which are dedicated to a particular modality (e.g., inertial, tactile, or ambient sensing). Our goal with the Stack was to move away from the large single circuit board developed for the Expressive Footwear and adopt a more compact and configurable geometry. Projects in which we have used the Stack include a pair of shoes designed for portable biomotion analysis and physical therapy via interactive music [10]. A single Stack layer measures 1.4"x1.4", about the size of a large wristwatch, but the large connectors required caused the Stack system to protrude significantly when several boards were used. In addition, the RF transceiver used at the time was capable of only 115 kbps. Although this is a factor of nearly 6 above the data rate of the Expressive Footwear board, it is still too small to enable a performance with multiple dancers. Assuming that each node transmits a 16-byte packet including a header, 6 12-bit inertial sensor values, and 3 16-bit capacitive or other analog measurements, two dancers wearing 4 nodes each should require 102.4 kbps. However, because of the time overhead associated with blocking reads/writes, it is difficult for a low latency protocol to achieve the full bandwidth. In practice (although this depends on the particulars of the implemented low-level RF protocols) achieving adequate reliability has restricted the system to 60% of the available bandwidth, thus pushing the requirements for two dancers up to 170.7 kbps.

Although compact sensor clusters have been developed at other institutes, none have the characteristics that we need for unobtrusively instrumenting a dance ensemble. The rich ancestry of Motes are quite established in sensor networks, but most support mainly peer-peer routing at lower data rates than needed here. Likewise, the Smart-Its and its descendants [11] are designed to work at data rates similar to the

Stack. Flety and collaborators at IRCAM [12] have built wireless sensor networks that use a similar transceiver as used in the Stack (hence also exhibit limited data rate) and others that use the WiFi 802.11 standard, which tends to be much too power hungry for efficient continuous operation with a modest battery. Emmanuel Tapia of the MIT Media Lab has designed very compact wireless accelerometer sensors capable of higher data rates [13], but our application requires more sensor degrees of freedom.

The design presented here includes a full six axis IMU, node-to-node capacitive proximity sensing, and flexible expansion capabilities, combined with a Nordic nRF2401 radio. The sensor node (Fig. 1) measures 1.65"x1.65"x0.5", not including the protruding antenna and external lithium polymer battery. It is thus slightly wider but much slimmer than the Stack. We chose to mount the RF transceiver on a daughter card, both to save board space and to achieve optimal RF performance. We also chose to decouple the battery from the main circuit board, so that it could be affixed to a strap rather than adding to the bulkiness of the sensor package. Screw mount holes are provided so that the node can easily be built onto a strap that secures it to the wrists or ankles.

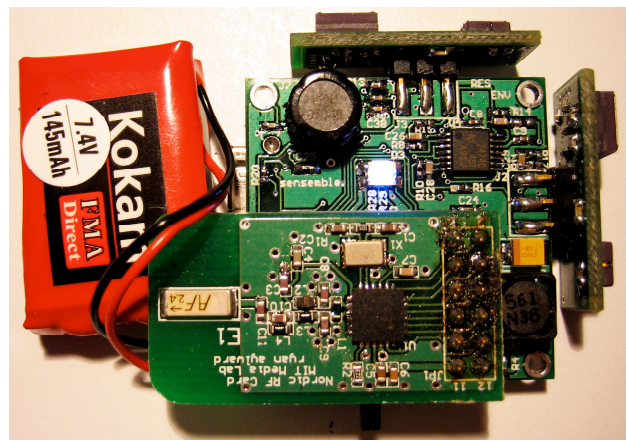


Figure 1. Sensor node and battery pack.

The nRF2401 data radio is a small, low-power, 2.4 GHz device providing up to 1Mbps data rates. Our communications protocol is a TDMA scheme [14] in which a basestation polls the network for data at the sampling rate, and each node responds within a preprogrammed time slot. The basestation then transmits the data to a central computer via USB for processing. Using this scheme, one basestation can handle full state updates at 100Hz for over 30 nodes. This is a significant bandwidth improvement over previous designs. The workable RF range on these devices appears to be on the order of 50 feet, depending on the local RF environment.

The IMU is made up of Analog Devices ADXRS300 rate gyros and ADXL203 accelerometers, as well as associated analog circuitry. Sensor signals are collected by the 12-bit analog to digital converter built into the

onboard processor, a TI MSP430. This microcontroller was favored because of its low power consumption, capable A/D, and ample I/O, as well as its use in several of our group’s ongoing projects.

The node-to-node capacitive proximity sensor operates by alternating transmit and receive modes on each of the sensor nodes, while the body is grounded – at any time, one node transmits while the others listen. During transmit mode, the microcontroller drives an LC oscillator, which generates a high amplitude (tens of volts) pulse at 90 kHz. During receive mode, this signal is picked up, amplified, and sampled in quadrature to determine its amplitude without the need for phase coherence. The transmit and receive modes on each node are able to share the same electrode very efficiently, thanks to an amplifier design inspired by Joshua Smith’s School of Fish [15]. The capacitive electrode can be integrated into the strap securing the sensor package to the body using conductive textiles such as Bekiweave [16].

Additional capabilities include 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 photoresistor. All of these optional signal lines are broken out to a compact expansion port, which also acts as the programming interface.

Power consumption is always of prime importance in the design of wireless sensors; the power source tends to be the largest and most cumbersome component of the system. Unfortunately, our desire to operate continuously with three rate gyros prevents this design from meeting traditional low-power requirements. Each gyro may consume up to 30mW, and their slow setup time prevents them from being power cycled. The data radio is also power hungry, but this can be managed in code by minimizing the amount of time spent in receive and transmit modes. Ultimately, we chose to operate the system with lithium polymer batteries because they are lightweight, compact, and rechargeable. With two 145mAh cells in series, as pictured above (Fig. 2), the node can operate for four hours on one charge.

4. Results

The major advantage of having large enough bandwidth to operate multiple sense points on multiple wearers simultaneously is the ability to obtain detailed information about the correlations present in group activity. In the context of a dance ensemble, time and spatial correlations can be used to determine which dancers are moving together, or perhaps moving complementarily in a call-and-response fashion. With this in mind, our preliminary analysis focuses mainly on the feasibility of extracting simple features that can be used to describe general group dynamics. The ability to detect

specific gesture patterns on an individual sensor node will be discussed below in future work.

The first task was to quantify both time separation and spatial similarity of gestures performed by multiple users. For simplicity, the network was limited to three sensor nodes, each worn on the right wrist of three test subjects. In a pair of tests, subjects were asked to raise and lower their right hands, first simultaneously and then in sequence. In the last test, subject one raised and lowered a hand as before, subject two performed a qualitatively “similar” but distinct gesture, and subject three performed a qualitatively “completely different” gesture.

To find the time separation between similar signals, cross-correlation is a natural choice. Here we have favored the similar measure of cross-covariance, because of the bias present in the raw inertial data. For each segment of test data, subject one was regarded as the reference and cross-covariance was calculated for the other subjects with respect to subject one. It can be seen that the location of maximum cross-covariance correctly estimates the time lag between similar gestures performed in succession (Fig. 3). As might be expected, cross-covariance also appears to be useful for determining the time delay between disparate gestures, but with diminishing accuracy.

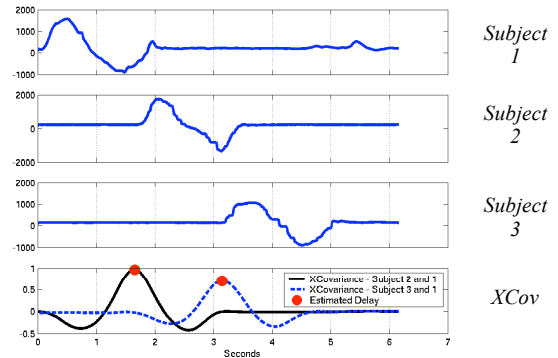


Figure 3. Raw pitch gyro data and resulting cross-covariance for hands raised and lowered in sequence.

The nature of cross-covariance as a signal-matching technique would suggest that the peak magnitude gives a measure of the strength of the correlation at the location of the peak. In the context of our sensor data this translates loosely into a measure of spatial similarity between gestures. Indeed, Figure 4 illustrates that as the disparity between gestures increases, the height of the peak cross-covariance decreases. This is a satisfying result in that a cross-covariance calculation can be used to determine both the time and spatial correlation of group movements.

Another useful group feature is the mean activity on each sensor axis. To determine this, the windowed variance of each sensor signal is calculated and then the results are averaged across subjects (Fig. 5). Windowed variance on individual sensor nodes can be used to

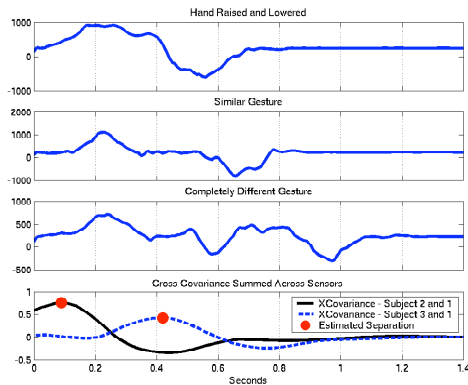


Figure 4. Raw pitch gyro data and resulting cross-covariance for one subject raising and lowering a hand, one subject performing a “similar” gesture, and one subject performing a “different” gesture.

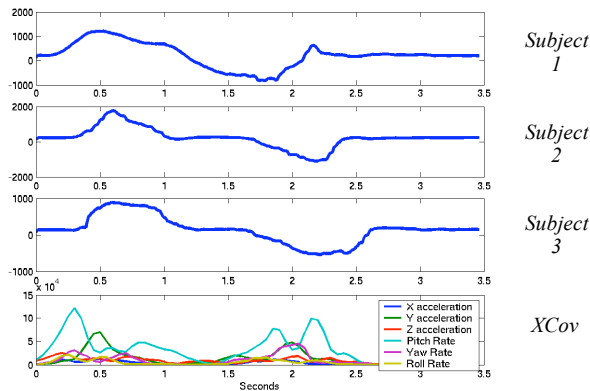


Figure 5. Raw pitch gyro data and average windowed variance across subjects as hands are raised and lowered in unison.

determine when gestures begin and end if there is enough pause between successive motions. Average windowed variance across the group extends beyond this to provide a measure of the predominant axes of collective motion during a period of activity. For instance, when most of the group is raising and lowering their hands, the average variance on the pitch gyro dominates (Fig. 5). This supports our intuition that raising and lowering the hand involves a rotation in pitch and little other activity. Extracting this information from average windowed variance may simplify the task of detecting specific or generic classes of gestures by determining which sensor signals are most important, or by defining a subgroup that is performing a similar gesture before applying heavier analytical techniques. One can also imagine a situation in which the correlation measurements discussed above are desired, but it is unclear who should be interpreted reasonably as a “reference” for the rest of the group. By comparing the average group variance to the individual variance, one can determine if the motions of one subject are characteristic of the entire group, or lie outside the norm.

5. Future Work

We plan to focus on extracting more sophisticated features from the data of individual nodes and training gesture models on these features, to investigate the possibility of classifying specific gestures being performed. Models for complex gestures could be based on a highly specific set of movements as defined by a choreographer, or based on sequences of smaller gesture units. In either case, group features as we have discussed here will greatly reduce the complexity of the problem, as it may no longer be necessary to detect a particular gesture being performed by every limb of every dancer.

6. References

- [1] di Perna, A., “Tapping into MIDI,” Keyboard Magazine (July 1988), p. 27.
- [2] Paradiso, J., et al, "Design and Implementation of Expressive Footwear," IBM Systems Journal, 39(3&4), October (2000), pp. 511-529.
- [3] Siegel, W. and Jacobsen, J., “The Challenges of Interactive Dance: An Overview and Case Study,” Computer Music Journal 22, No. 4, 29-43 (1998).
- [4] Coniglio, M., The MidiDancer system, see: <http://www.troikaranch.org/mididancer.html>.
- [5] See: <http://www.measurand.com/>
- [6] Zacks, R. (1999). Dances with Machines. Technology Review, (May/June), pp. 58-62.
- [7] Camurri, A., S. Hashimoto, M. Ricchetti, R. Trocca, K. Suzuki, and G. Volpe. “EyesWeb – Towards Gesture and Affect Recognition in Dance/Music Interactive Systems,” Computer Music Journal, 24(1), pp. 57-69.
- [8] Wechsler, R., Weiss, F., and Dowling, P. “EyeCon – A motion sensing tool for creating interactive dance, music, and video projections,” Proc. of the Society for the Study of Artificial Intelligence and the Simulation of Behavior (SSAISB) Convention, Leeds England, March 29, 2004.
- [9] Benbasat A.Y. and Paradiso, J.A., “A Compact Modular Wireless Sensor Platform,” in Proc. of Information Processing in Sensor Networks (IPSN), Los Angeles, Apr. 25-27, 2005, pp. 410-415.
- [10] Paradiso, J.A., et al., Interactive Therapy with Instrumented Footwear, in the Proc. of the ACM Conference on Human Factors and Computing Systems (CHI 2004), Extended Abstracts, Vienna, Austria, April 27-29, 2004, pp. 1341-1343.
- [11] Holmquist, L.E., et al, “Building intelligent environments with Smart-Its,” IEEE Computer Graphics and Applications Magazine, vol. 24, no. 1, 2004, pp. 56-63.
- [12] Flety, E., “The WiSe Box: a Multi-performer Wireless Sensor Interface using WiFi and OSC,” in the Proc. of NIME 05, Vancouver CA, May 26-28, 2005, pp. 266-267.
- [13] Munguia Tapia, E., et al, "MITes: Wireless portable sensors for studying behavior," in Proceedings of Extended Abstracts Ubicomp 2004: Ubiquitous Computing, 2004.
- [14] Lovell, S.D., A System for Real-Time Gesture Recognition and Classification of Coordinated Motion, M. Eng. Thesis, MIT EECS Dept., Jan. 2005.
- [15] Smith, J.R., Electric Field Imaging, Ph.D. Thesis, MIT Media Lab, February 1999.
- [16] See: <http://www.bekaert.com>