# Wireless sensor interface and gesture-follower for music pedagogy

Frederic Bevilacqua, Fabrice Guédy, Norbert Schnell
Real Time Musical Interactions
Ircam - CNRS STMS
1 place Igor Stravinsky
75004 Paris – France
+ 33 1 44 78 48 31
{Frederic.Bevilacqua, Fabrice.Guedy, Norbert.Schnell}@ircam.fr

Emmanuel Fléty, Nicolas Leroy
Performing Arts Research Team
Ircam Centre Pompidou
1 place Igor Stravinsky
75004 Paris – France
+ 33 1 44 78 15 49
{Emmanuel.Flety, Nicolas.Leroy}@ircam.fr

## ABSTRACT
We present in this paper a complete gestural interface built to support music pedagogy. The development of this prototype concerned both hardware and software components: a small wireless sensor interface including accelerometers and gyroscopes, and an analysis system enabling gesture following and recognition. A first set of experiments was conducted with teenagers in a music theory class. The preliminary results were encouraging concerning the suitability of these developments in music education.

## Keywords
Technology-enhanced learning, music pedagogy, wireless interface, gesture-follower, gesture recognition

## 1.1 INTRODUCTION
The recent developments in the fields of movement analysis and gesture capture technology create appealing opportunities for music pedagogy. For example, traditional instruments can be augmented to provide control over digital musical processes, altering standard instrument practice and offering potentially complementary pedagogical tools. Moreover, the development of novel electronic interfaces/instruments generates even more different paradigms of music performance, giving rise to potential novel approaches in music education.

In this article we present a gestural interface that was integrated in a music education context. Both hardware and software components were developed and are described here. First, we report on the design of a relatively inexpensive miniature wireless sensor system that is used with accelerometers and gyroscopes. Second, we describe a gesture analysis system programmed in the Max/MSP environment to perform gesture recognition and following. The complete prototype enables us to experiment with various pedagogical scenarios. This research is currently conducted in the framework of the European I-MAESTRO project on technology enhanced learning, focusing on music education [23].



The motivation for this work is grounded in our pedagogical approach that considers physical gesture [11] as a central element for performance but also for the embodiment of music concepts and theory. Our working hypothesis is that specific gestrual interactive systems can enhance this pedagogical approach. Even if similar or complementary tools have been already proposed and carried out [7][9][10][13][14][27], the use of digital technology and gestural interfaces in music pedagogy is at its very beginning. Any use of new technology in music education represents difficult challenges, nevertheless we believe that such an approach offers great potential.

This paper is divided in three separate parts. The first two parts concern the technological developments, respectively the wireless sensor interface and the gesture follower/recognizer. In the third part, we present the pedagogical scenarios and the preliminary results we obtained after a first set of trials in music classes.

## 2. WIRELESS INTERFACE AND SENSORS

### 2.1 Requirements
We developed and reported previously on several wireless interfaces, that were used in applications including the *augmented violin* project [3] and dance performances [8]. The experience we gained with these applications helped us to define requirements for the interface presented here.

Precisely, we developed in 2005 an 802.11b WiFi portable acquisition device called the *WiSe Box* [8]. An important advantage resides in the possibility of working simultaneously with multiple devices thanks to the different WiFi channels. The device offers 16 sensor channels, sampled on 16-bit resolution at 200Hz sampling rate, for overall dimensions of 110×65×28 mm.

The aim of the development described here was to maintain most of the specifications of the *WiSe Box* while drastically reducing its

size and power consumption. The following requirements were used as guidelines:

- compact size and weight enabling the sensors, wireless transmitter and battery to fit in a light handheld device

- low power consumption, autonomy for standard rehearsal and performances

- simultaneous use of multiple devices

- low latency and sufficient accuracy for music performance (typically sampled at 200 Hz on 10 bits).

- cost effective

- limited expertise and skills to operate the device. Robustness and reliability compatible within a pedagogical context.

## 2.2 Related works

Similar interfaces were reported recently. The company *Infusion Systems* proposes the Wi-microDig, a Bluetooth sensor interface [21]. The *CrossBow* company has a product line of MICA modules designed for sensor nodes to be spread in various distant locations in large space [28]. Paradiso and coworkers at the MIT MediaLab developed a wireless and compact multi-user sensor system for dance performance featuring highly reduced size, high-end electronics and low-power supply solutions [1]. Chou et al. developed at the University of California Irvine a thumb-sized sensor interface focused on node spreading [20].

## 2.3 Data acquisition and transmission

We based our design on the XBee from *MaxStream* [24], which is a small form factor OEM module with simple communication means operating with the 802.15.4 IEEE standard. This standard, also known as *Zigbee,* is a variation of the 802.11 standard designed for embedded and low power wireless electronic devices. Most of the important features of a wireless network architecture are available: unique MAC address to identify each transceiver, beacon frames for node discovery and to wake up units in sleep mode, Carrier Sensing Multiple Access with Collision Avoidance (CSMA/CA) to share the bandwidth of a single frequency channel, and finally multiple channels over the ISM band.


**Figure 1. Right: the module with Xbee, battery and sensors (on the back of the module). Left: receptor, showing the Ethernet and power supply connectors.**

The XBee modules allow for the use of basic RS-232 wireless serial links to high-speed sensor networks. Each device can be considered as a serial modem (115200 bauds) and embeds a microcontroller that responds to AT commands for configuration and data transmission/reception.

We used a specific version of the XBee module firmware that includes its own microcontroller operating the wireless section. The device features 6 analog inputs with a 10-bit AD converter. Thus, there is no need to use any additional microcontroller, and direct wiring of analog sensors to the XBee module is possible.

The CSMA/CA protocol allows for several transceivers to be merged into a single master but reducing the maximum data rate of each digitizer. To guarantee the highest data transmission performance, an individual receiver must be used for each emitter.

## 2.4 Data reception and computer communication

The sensor data are sent as Open Sound Control messages [25] over UDP, as often found in recent sensor digitizers such as the *WiSe Box*[8], *Toaster*, the *Kroonde* [6] or the *Gluion* [22]. The wireless receiver device uses a paired XBee module communicating with a PIC18F4520 Microchip microcontroller. The micro-controller communicates with the Microchip Ethernet controller ENC28J60 through a Serial Peripheral Interface (SPI) synchronous serial link. This enables the transmission of UDP packets with the OSC protocol. To reduce size and cost, we used a specific RJ-45 port containing both the link/data LEDs and the Ethernet isolation transformer. The data is received and processes by Max/MSP on the host computer connected to the receiver module.

## 2.5 Sensors and power supply

We choose a 5D Inertial Measuring Unit sensor including a ± 3g three-dimensional accelerometer from Analog Device (ADXL330) and an *Invensense* IDG-300 dual-axis gyroscope. Those two parts are available from *Spark Fun Electronics*[26] assembled on a 22x20 mm Printed Circuit Board.

For power supply, we chose a lithium-polymer flat battery of 3.6 volts / 140 mAh (30x20 mm). This very small form factor allowed us to slide the battery between the main PCB and the XBee module, making the whole wireless module to fit a volume of 38×27×11 mm. The overall device consumes 40 mA @ 2.8 volts and has an autonomy of 3 hours. If weight and size are not critical, bigger battery might be used: we tested a 340 mAh model that last 7 hours. Note that one of the analog input can be used to monitor the battery voltage.

## 2.6 Performances and applications

The sensor stream is digitized (10bit) and transmitted at a framerate of 200 Hz for each emitter/receiver pair. Several modules can operate simultaneously. The range of the transmitter is 10 meters in an open area. This range may appear limited compared to usual OEM single RF frequency modules, but this does not represent a constraint for our applications where the receivers can be placed close enough to the emitter. If needed, larger range can be achieved by using the PRO version of the XBee modules, although autonomy might be reduced in such a case.

The wireless sensor interface was fully tested and used in two applications: handheld devices for free gesture interaction and augmented string instruments (string quartet for example). This article concerns the first type of applications.

# 3. REALTIME GESTURE ANALYSIS

## 3.1 Gesture following/recognition

The development of the *gesture-follower* is pursued with the general goal to compare in real-time a performed gesture with a set of prerecorded examples, using machine learning techniques. Similar approaches have been reported [6][12][13][14][15][16] and are often used in implicit mapping strategies.

In our context, a "gesture" is defined by its numerical representation produced by the capture device. Technically, this corresponds to a multidimensional data stream, which can be stored in a matrix (e.g. row corresponding to time index, and column to sensor parameters). A multimodal ensemble of temporal curves can be directly accommodated within this framework, as long as all curves have identical sampling rate.

### *3.1.1 Following*

The *gesture-follower* indicates, during the performance, the time location (or index) in relation to the recorded references. In other words, the *gesture-follower* allows for the real-time alignment of a performed gesture with a prerecorded gesture.

Figure 2 illustrates the computation, performed each time a new data is received, of the corresponding the time index of the reference. This operation can be seen as a real-time time warping of the performed gesture to the recorded reference.

### *3.1.2 Comparing and recognizing*

The process explained in the previous section can be performed with several references simultaneously. In this case, the system compute also the likelihood values for each reference to match the performed gesture. An example of this process is illustrated in Figure 3 where the performed gesture is compared to two other examples.

As shown in Figure 3 the likelihood values are updated continuously while the performed gesture is unfolding. The result of the recognition can therefore vary from the beginning, middle or the end of the performed gesture. Gesture recognition can be achieved by simply selecting the highest likelihood, at a chosen time.

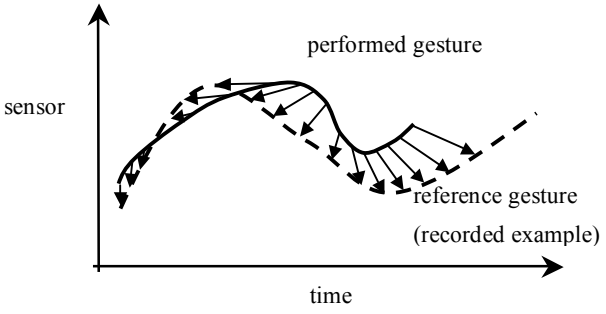


**Figure 2. The *following* paradigm: the performed gesture is time warped to a given reference.**

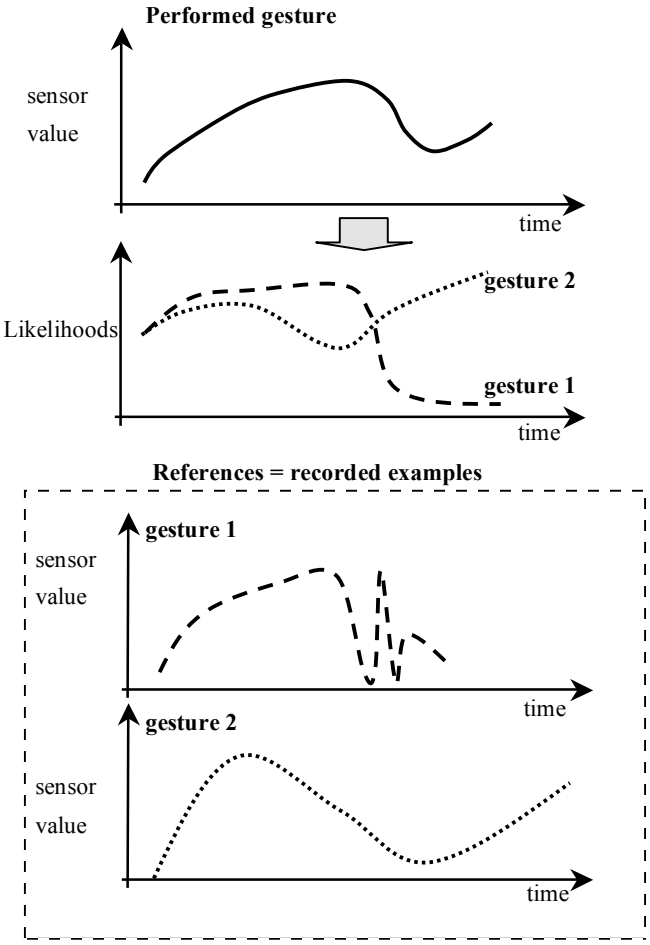


**Figure 3. Comparison and recognition paradigm.**

## 3.2 Algorithm

The two paradigms we described above, *following* and *recognition* can be directly implemented using Hidden Markov Models (HMM) [19]. Generally, the parameters of Markov models are estimated using the Baum-Welch algorithm using a large set of examples. In our case, we choose a simplified learning method enabling the use of a single example to determine the model parameter. To achieve this, assumptions are made on the expected variations within a class of gesture. This procedure can lead to a suboptimal determination of the Markov Model parameters. However, the possibility of using only a single example represents a significant advantage in term of usage.

### *3.2.1 Learning*

The learning process is illustrated in Figure 4 where the temporal curve is modeled as left-to-right Markov chain. The learning example is first downsampled, typically by a factor 2, and each sample value is associated to a state of the Markov chain. Assuming a constant sampling rate, the left-to-right transition probabilities are constant and directly related to the downsampling factor. For example, in the case of downsampling of factor *n*, the transition probabilities are equal to *1/n,* ensuring the Markov chain to model adequately the temporal behavior of the learning example.

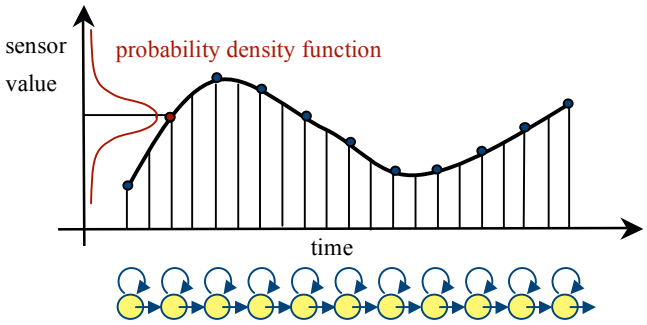


**Figure 4 Learning procedure: a left-to-right HMM is used to model the example, downsampled by a factor 2.**

The observation probability function for each state is considered to be a multidimensional Gaussian model with a mean $\mu_i$ and a variance and $\sigma^2_i$, where $i$ is the state number. The mean $\mu_i$ is set to the value of the recorded gesture. The variance value is a factor adjusted by the user, which must match approximately expected variations between the performed and recorded gestures. In most of our experiments we found that the variance value is not critical since the recognition is based on a comparison process.

### *3.2.2 Decoding*

Consider the performed gesture as a partial observation sequence $O_1..O_T$, corresponding to the performed gesture values from time *1* to *T* (sample index). The probability $\alpha_t(i)$ of this partial sequence and state *i* is computed from the standard forward procedure in HMM [19].

The *following* procedure corresponds to determine the most likely state *i,* denoted $j(t)$, for all time *1...t*:

$$j(t) = \arg\max_i \left[\alpha_t(i)\right] \quad 1 < t < T \qquad \text{Eq. 1}$$

Since the Markov chain has a simple left-to-right structure, the computed sequence of $j(t)$ reports time indexes of the time-warped sequence to the learned example (as shown in Figure 2).

The *comparison* and *recognition* procedure corresponds to compute the likelihood of the observation sequence for a given example (i.e. Markov model)

$$likelihood(example) = \sum_i \alpha_t(i) \qquad \text{Eq. 2.}$$

### *3.2.3 Implementation*

The *gesture-follower* is implemented as a set of Max/MSP modules integrated in the toolbox MnM [2] of the library FTM (LGPL licence) [18]. It takes advantages of the data structure of FTM for Max/MSP such as matrices and dictionaries. An example is freely available in the FTM package, under MnM/example.

## 4. PEDAGOGICAL EXPERIMENTS

Conducting is an important part of musical education for all instrument players. It is an essential part of practice training, closely related to music theory. While teaching methods for small children or beginners are often based on playful approaches and exercises focusing on body movements (e.g. Dalcroze, Menuhin), music education at higher levels tends to underestimate these aspects. For some mid-level students, this may lead to a rigid posture and stiff gestures in their instrument practice.

**Figure 5. Teacher and student using the system during a music class. The teacher holds the wireless sensor module during the learning phase.**

We performed two experiments with students during a regular music theory lesson (music school "Atelier des Feuillantines" in Paris). Minimum perturbation was sought: the lesson was conducted by the usual teacher and following the usual lesson structure (Figure 5).

The pedagogical aim of the exercise was to experience and practice "smoothness" and "fluidity" of musical gestures. The prototype was used to continuously synchronise a chosen soundfile to a conducting gesture performed with the wireless module. The teacher starts the exercise by recording the reference gesture: he conducts while listening to the soundfile. In a second phase, the students use the system to "conduct" the music, as further explained in the next section.

### 4.1 Interaction paradigm

The gesture-follower was used to control the playback of soundfiles. The time index output by the *gesture-follower* directly determines the position in the soundfile. Two types of time-stretching were used: granular synthesis or phase vocoder implemented with the Gabor library of FTM [17][18].

On a practical level, the procedure is as follows:

1. Record mode: Record the gesture example while listening to the sound file. This step provides a gesture example that is synchronized with the soundfile.

2. Play mode: The soundfile playback speed varies according to the *gesture-follower* output, depending of the temporal variation in the gesture performance.

Separate soundfiles can be associated to different recorded examples. Different playback schemes are possible. First, the *recognition* feature can be used for the selection of the soundfile corresponding to the most likely gesture. Second, the different soundfiles can be played back simultaneously, and mixed according to the likelihood values.

This interaction paradigm can be used to simulate orchestral conducting. Similar applications have been proposed and implemented by several groups [5][13][14]. However, our approach is distinct from those on various points.

First, the gesture is considered here as a continuous process. In particular, no beat detection is used. This point has important consequences discussed in the next section.

Second, the choice of the gesture is totally open and can be chosen with a very simple and effective procedure. As mentioned earlier, a single recording of the gesture is sufficient to operate the system. This flexibility allows us to elaborate pedagogical scenarios where the conducting pattern can be freely chosen and adjusted by the user. This point is further developed in section 4.3.

## 4.2 Experiment 1: Conducting

After starting the software in "record mode", the teacher records a usual beat pattern gesture while listening to an excerpt of the soundfile. For example an excerpt of the Rite of Spring was chosen for its changes of metric.

The software is then switched in "follow" mode and the students are asked to use the system to "conduct" the soundfile. An excerpt of a recorded beat pattern and the time-warped performed gesture is shown in Figure 6.

Since the system does track the entire gesture and not only the beats, the gesture between beats is important and affect directly the conducting procedure. Therefore, the audio playback speed depends directly on the overall movement quality. For example, if

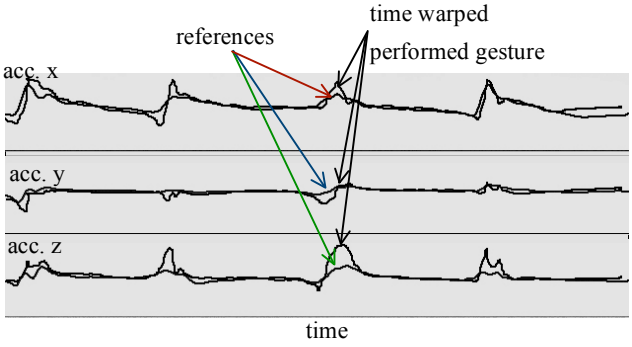


**Figure 6. 4-beat gesture as recorded by the 3D accelerometer**

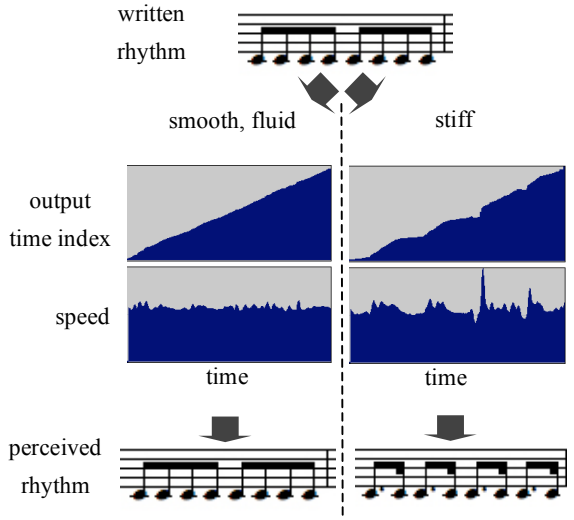


**Figure 7. Effect of smoothness and fluidity in the performance of the 4-beat conducting pattern.**

the student gesture does not match the smoothness and fluidity of the teacher gesture, a striking modification of the rhythmic pattern of the conducted sound appears (Figure 7). This effect provides a direct sonic feedback to the students of its overall gesture quality, who can then progressively learn, "by ear", how to perform a smooth and fluid gesture.

## 4.3 Experiment 2: Free gesture exploration

In this experiment, the students were asked to find a free gesture they felt appropriate to various soundfiles. Various gestures were experimented by the students to control the temporal flow of music/sound.

Different cases were tested, including the excerpts used for experiment 1. After experiencing traditional beating patterns, the students were able to try other types of gesture than usual conducting gestures. Voice recordings of the students were also used. The association of a free gesture to a voice recording allowed them for instance to alter the rhythm/prosody.

## 4.4 Discussion and further work

The experiments reported here must be understood as exploratory, and any definite conclusion should be avoided at this early stage. Importantly, the approach proposed here should be understood as complementary to traditional music teaching (rather than a replacement). The system was first tested during regular music theory lessons and the students were highly motivated by the experiments. Moreover, they immediately pointed out its creative potential. The teacher felt significant improvements of student awareness to key aspects of performance practice, for example musical phrasing. We summarize below important points that the experiments brought out, defining interesting paths for future work.

### *4.4.1 Smoothness and fluidity*

To experience smoothness and fluidity in a musical context was one of the goals of experiment 1. The control of these "gesture qualities" is crucial in music performance and interpretation, and represents generally a difficulty for young students. As a matter of fact, a usual problem among beginners (typically older than 10 years) resides in their overall body rigidity; they tend to move only the body parts touching the instrument. We found that our system was an interesting approach to stimulate adequate motion. Further experiments will concern attaching sensors to different body parts.

### *4.4.2 Breathing*

Breathing is a well-known issue in music practice (directly linked to the point previously discussed). For example, students practicing "mechanically" in a stiff position tend to play often in apnoea, blocking their breathing. These moments of apnea are evidence of insufficient connections between breathing and playing. The two experiments suggest that the system could be used to sonify particular gesture aspects directly linked to breathing and therefore helping the practice of musical phrasing/breathing.

### *4.4.3 Link between intention and gesture*

The understanding of the musical structure and other compositional aspects of a musical piece (cadences for instance) usually help music interpretation and expression. Lack of theory understanding prevent students to elaborate consistent music interpretation. Our approach can potentially give opportunities to to experience in an interactive way some aspects of music theory.

## 5. CONCLUSION

We presented a set of hardware and software tools that were integrated in a fully functional prototype. On the technological side, the developments were found to be robust, and allowed for rapid prototyping of pedagogical experiments. A single person was able to install and operate the system seamlessly.

The first use of the system in a music class was encouraging and allowed us to confirm our approach. Larger scale experiments are currently planned with additional sound processing possibilities, including various sound synthesis modules. The same wireless sensor system and the gesture-follower are currently adapted to the case of violin playing.

## 6. ACKNOWLEDGMENTS

The I-MAESTRO project is partially supported by the European Community under the Information Society Technologies (IST) priority of the 6th Framework Programme for R&D (IST-026883, www.i-maestro.org). Thanks to all I-MAESTRO project partners and participants, for their interests, contributions and collaborations.

We would like to thank Remy Muller, Alice Daquet, Nicolas Rasamimanana, Riccardo Borghesi, Diemo Schwarz and Donald Glowinski for contributions to this work and fruitful discussions.

## 7. REFERENCES

[1] Aylward R. and Paradiso J., "Sensemble: Awireless, compact, multi-user sensor system for interactive dance" *Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06)*, Paris, France, 2006.

[2] Bevilacqua, F., Muller, R., Schnell N. " MnM: a Max/MSP mapping toolbox ", *Proc. of the International Conference on New Interfaces for Musical Expression (NIME 05),* Vancouver, Canada, 2005.

[3] Bevilacqua F., Rasamimanana N., Fléty E., Lemouton S., Baschet F., "The augmented violin project: research, composition and performance report", *Proc.s of the International Conference on New Interfaces for Musical Expression (NIME 06),* Paris, France, 2006.

[4] Benbasat, A. Y., and Paradiso, J. A., "An Inertial Measurement Framework for Gesture Recognition and Applications" In Ipke Wachsmuth, Timo Sowa (Eds.), "*Gesture and Sign Language in Human-Computer Interaction*" International Gesture Workshop, GW 2001, London, UK, 2001 Proceedings, Springer-Verlag, Berlin, 2002, pp. 9-20.

[5] Borchers J., Hadjakos A., and Mühlhäuser M., "MICON: A Music Stand for Interactive Conducting", *Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06),* pp 254-259, Paris, France, 2006.

[6] Coduys, T., Henry, C. and Cont, A. "TOASTER and KROONDE: High-Resolution and High-Speed Real-time Sensor Interfaces" *Proc. of the International Conference on New Interfaces for Musical Expression (NIME-04),* Hamamatsu, Japan, 2004.

[7] Ferguson F., "Learning Musical Instrument Skills Through Interactive Sonification" *Proc. of the International Conference on New Interfaces for Musical Expression (NIME 06)*, pp 384-389, Paris, France, 2006.

[8] Flety, E., "The WiSe Box: a Multi-performer Wireless. Sensor Interface using WiFi and OSC", *Proc. of the International Conference on New Interfaces for Musical Expression (NIME05),* Vancouver, Canada, 2005.

[9] Guédy, F, "Le traitement du son en pédagogie musicale", L'Inouï, Vol. 2, Ircam – Editions Léo Scheer.

[10] Puig V., Guédy F., Fingerhut M., Serrière F., Bresson J., Zeller O. "Musique Lab 2: A Three Level Approach for Music Education at School", *Proc. of the International Computer Music Conference* (ICMC 2005*)*, Barcelona, Spain, 2005.

[11] Iazzetta, F., "Meaning in Music Gesture", Trends in gestural Control of Music, Marc Battier & Marcelo M. Wanderley (Eds.) Paris-IRCAM - Centre Pompidou - CD-ROM, 2000.

[12] Kolesnik P. and Wanderley M. M.,"Recognition, Analysis and Performance with Expressive Conducting Gestures.", *Proc. of the 2004 International Computer Music Conference (ICMC 2004),* Miami, USA, 2004.

[13] Lee E., Grüll I., Kiel H., and Borchers J., "3conga: A Framework for Adaptive Conducting Gesture Analysis" *Proc. of the International Conference on New Interfaces for Musical Expression (NIME06),* pp. 260-265, Paris, 2006.

[14] Lee E., Karrer, T. and Borchers J.. "Toward a Framework for Interactive Systems to Conduct Digital Audio and Video Streams" Computer Music Journal, 30(1):21-36, 2006.

[15] Merrill, D. and Paradiso, J.A Personalization, "Expressivity, and Learnability of an Implicit Mapping Strategy for Physical Interfaces*", Proc. of CHI 2005 Conference on Human Factors in Computing Systems*, Extended Abstracts, ACM Press, Portland, OR, April 2-7, 2005, pp. 2152-2161.

[16] Pritchard B., Fels S.: "GRASSP: Gesturally-Realized Audio, Speech and Song Performance". *Proc. of the International Conference on New Interfaces for Musical Expression (NIME06)*, 272-271, Paris, France, 2006.

[17] Schnell N., Schwarz, D. "Gabor, Multi-Representation Real-Time Analysis/Synthesis", *COST-G6 Conference on Digital Audio Effects (DAFx)*, Madrid, Spain, 2005.

[18] Schnell, N., Borghesi, R., Schwarz D., Bevilacqua, F. Müller, R. " FTM — Complex data structures for Max ", *International Computer Music Conference (ICMC),* Barcelona, Spain, 2005.

[19] Rabiner L. R., Juang B., Tutorial on hidden Markov models and selected applications in speech recognition, *Proc. of the IEEE*, vol. 77, no. 2, pp. 257-285, 1989.

[20] http://ecomote.net/

[21] http://www.infusionsystems.com/

[22] http://www.glui.de/

[23] http://www.i-maestro.org/

[24] http://www.maxstream.net/

[25] http://www.opensoundcontrol.org/

[26] http://www.sparkfun.com/

[27] http://www.toysymphony.net/

[28] http://www.xbow.com