# DATA ANTICIPATION FOR GESTURE RECOGNITION IN THE AIR PERCUSSION

*Vincent Goudard, Christophe Havel, Sylvain Marchand,*
*and Myriam Desainte-Catherine*
SCRIME – LaBRI, University of Bordeaux 1, 351 cours de la Libération,
F-33405 Talence cedex, France
www.scrime.u-bordeaux.fr, www.labri.fr

## ABSTRACT

The use of computers as Digital Musical Instruments (DMI) rises many problems due to the separation between gesture and sound production. However, computational power allows now for complex real-time analysis and gesture anticipation and recognition. We present here an interactive performance system for a percussionist playing in the air, with no physical percussion. Since the gesture sensors are not perfect, we have to deal with unreliable data. And since we expect a real-time sound synthesis synchronous with the strikes of the percussionist, we have to anticipate his gesture to some extent, in order to forecast his strikes, because we cannot wait for the strikes to occur without degrading the auditory feedback. We show that the use of linear prediction can help to both correct the data from the gesture sensors and anticipate the gesture of the percussionist. The difficult problem of gesture recognition is also discussed in the context of the "air percussion" project.

## 1. INTRODUCTION

If the expressiveness of an instrumentalist lies in his ability to play an instrument through an intuitive coordination of his gesture, in response to an intimate musical feeling and listening-understanding, one has to consider the long experience of traditional acoustic instrument players.

At the present time, our choice is limited to a restricted set of digital instrument interface imitating / modeling their acoustic equivalent, the most common being the MIDI keyboard for piano and organ. In particular, the percussion finds no suitable equivalent in the electronic instrumentarium. Pads almost reduce the sharpness of the play to a simple triggered event with velocity information, while the contact can take plenty of other subtleties (e.g. depending on the point of strike on the percussion, the kind of contact, etc.).

Furthermore, the contact is only a very small part of the whole gesture. In the artistic approach presented here, we choose to remove the contact and focus on the movements of the sticks in the air.

A few projects have been led until now to alleviate this problem, but technology was hardly able to render both static (position) and dynamic (movement) information. From Boie and Matthews "Radio-Baton" [12] covering a reduced surface, to accelerometer solutions that do not send position back [11], to video-based systems [1]

---

[1] http://www.gmem.org

which suffer from high computational load, the complex gesture of the percussionist remains with no clear modeling.

If the digital interface takes into account the ergonomic of the traditional instrumentarium, the composer is then in a position to define a new sonic environment, with unexplored sounds and interactions, while keeping the articulation expertise of the traditional instrumentalists.

After a brief presentation of the "air percussion" project in Section 2, we show in Section 3 some particularities of this project due to the behavior of the percussionist and give in 4 some interesting solutions for gesture data correction and anticipation based on a linear prediction technique. Finally, the difficult problem of gesture recognition is discussed in Section 5.

## 2. THE "AIR PERCUSSION"

### 2.1. Artistic Project

Since 2000 the second author (Havel, composer) develops "metamorphoses", a long-term project of investigation and experimental creation with the ambition of defining, through a succession of works performed in public, a process of musical writing for a group of chamber music performers.

The setting is as following: to put the instrumentalists in front of an electronic set-up with the same ergonomics that their custom acoustical instruments but with sounds and control completely different. In this case, the musicians must play according to the reactions of the electronic instrument and the musical propositions of the other musicians inside a formal structure proposed by the composer.

The electronic instrument consists of a synthesis module associated to a set of interrelations (mapping) between the gestures sensors and the control of the sound parameters. Each musician generally has many instruments (synthesis modules + mapping) at one's disposal, instruments he could select during the performance.

### 2.2. System architecture

In the context of this musical research, the SCRIME (LaBRI, University of Bordeaux 1) developed a scientific research project for the recognition of the percussionist gesture.

An air-percussion device thus was implemented from the obtained results, including a system of sensors connected to a computer which analyses the gestures of the
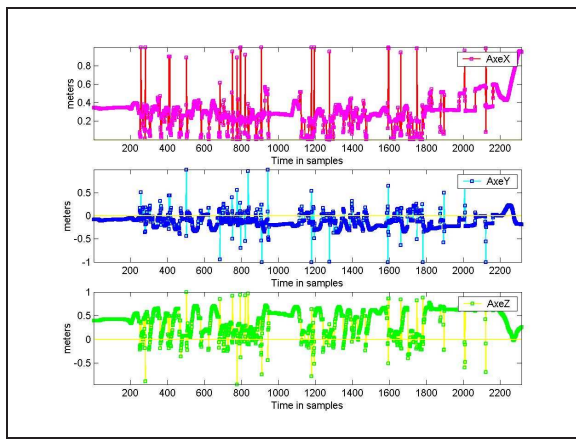
**Figure 1**. The FOB may deliver very disturbed signals depending on the electromagnetic environment.

percussionist, a system of graphic edition of the instruments and visualization of the trajectories of the strikes, and modules of sound synthesis.

The "air percussion" is a DMI using the "flocks of birds" [2] (FOB) as sensors put at the tip of drum sticks, receiving electromagnetic waves produced by a cubic emitter, and sending position and orientation information back to a computer.

This system was already presented in [1]. The particularity of this system is that the percussionist plays in the air, with no physical support, virtual objects that are placed in the space around him, relying with his traditional percussionist knowledge.

There are few systems that allow getting absolute position in space with precision at a reasonable price. May the system use ultrasonic, infrared, or electromagnetic technology (which is used in the FOB), we often face serious perturbations of the signal, especially in a performance context.

The "air percussion" system is made up of a PC running a program using the libflock, a dedicated library for accessing the hardware, and sending flocks events to a remote host by using UDP datagram that comply to CN-MAT's OpenSound control protocol [10]. These messages are then received by a Macintosh computer running Max / MSP for the sound synthesis.

The FOB is very sensitive to electromagnetic perturbations, thus delivering many erroneous data (cf. Figure 1). Gesture signal analysis and correction is performed on the PC, and is partly described later in this paper.

## 3. PERCUSSIONIST BEHAVIOR

### 3.1. Modeling the Percussionist Gesture

The model we used here makes use of a mapping layer based on a formalism proposed by the percussionist Dupin to teach percussion, and consisting in four different types of strikes, called: *down*, *up*, *piston*, and *muffled*.

These basic strikes can then be sequenced in various manners to form the following figures:

**roll** : the two sticks alternate the cycle [down, up] symmetrically;

---

[2] http://ascension-tech.com

**single roll** : the two sticks alternate pistons symmetrically;

**flam** : the two sticks alternate the cycle [down, up] asymmetrically, starting on an up.

Now, the movements of a percussionist in a real performance context are very complex, as they do involve not only the arms, but also the whole body. And just like speech is not a series of juxtaposed phonemes, the percussionist play is made of gesture, which can yet be identified as we pretend, but greatly depend on past and future movement. This had not been taken into account in previous studies, where the various types of strikes were considered independently.

In the peculiar case of the percussion, the gesture recognition consists not only in determining the shape of the trajectory but also in localizing the precise absolute space and time position of the strikes, to know which instrument (and which part of the instrument) was played. Another interesting parameter is the velocity of the gesture as well as the acceleration that permits to determine the shock point.

Another specific aspect of the air-percussion is that the musician doesn't have any tactile feedback as in the traditional instrument. Although it prevents the percussionist from playing successions of strikes using the skin bounce, the experience shows that it doesn't prevent the percussionist from playing while conserving most of his skills. Indeed, while the movement of the tip of the stick is affected by the skin bounce, the strike gesture is quite independent of the material [3] . On the other hand the auditory feedback plays a fundamental role.

### 3.2. Influence of Auditory Feedback

The first phenomenon we notice was that the percussionist would modify his gesture to get a sound that matches his expectation. Indeed, the "air percussion" still suffers from bugs at the present time, and intended strike do not always sound, while unwanted strikes are sometimes being triggered.

The percussionist thus intuitively adapts his gesture to compensate the system drawbacks, by trying to find the right zone of the space, where he gets the result he expects. One should hence be careful when analyzing such gesture to consider the gesture deformation with respect to sound output. Another point is that the auditory feedback appeared most useful in fast gestures like roll, for which the percussionist seems to use more the auditory than kinesthetic feedback to count strikes.

### 3.3. Accompanist Gesture

We also notice gestures that look similar to strikes that are not meant to sound, but rather to keep the tempo and support oneself on the beat. These gestures are often performed with a breath-in for fast traits like roll, as if the instrumentalist is preparing to dive (cf. Figure 2). We could also notice than in such musical parts where virtuosity is required, the body tends to find a stable position, with the head acting as a "stabilized inertial guidance platform" [9].

---

[3] Indeed, the instrumentalist DO NOT try to go through the skin (!): the return of the stick is not a consequence of the bounce, but performed on purpose.
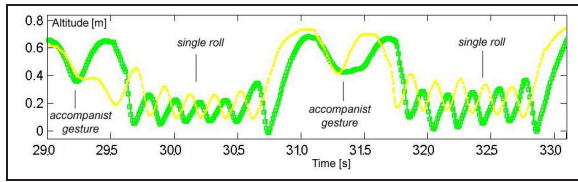
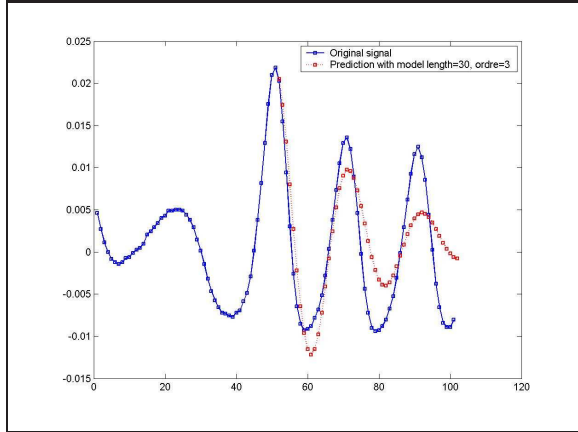**Figure 2**. Accompanist gestures: keeping the tempo (one curve for each hand).



**Figure 3**. Prediction of a *roll*, LPC coefficients computed with Burg's method (blue solid line: original data, red dotted line: prediction).

## 4. DATA CORRECTION AND ANTICIPATION

### 4.1. Using the Burg Method

To correct the FOB erroneous data and anticipate impacts exact time, we implemented a linear prediction coding (LPC) algorithm, using Burg's method for computing the coefficients. This method has already been used with success for extrapolating audio signals [3], and tracking the evolutions of the partials in the context of sinusoidal modeling [2].

Among the various existing methods to compute LPC coefficients, the Burg method was chosen due to its ability to generate synthetics signals close to natural ones (cf. Figure 3). Furthermore, the Burg method leads to minimum-phase filters, ensuring a fast and stable response, which is not the case with the autocorrelation or covariance methods. The Burg method is presented on both theoretical and computational points of view for example in [4].

### 4.2. Model Calibration

As far as we are concerned, we needed to find the filter order, and the model size, which were chosen after both theoretical and empirical considerations. In our case, since there is no shock of the sticks, we are observing gesture signals with exponential, or pseudo-sinusoidal variations. Hence, the filter order should be more than 2, but a too high order would produce undesirable oscillations. The model size (i.e. the time-window used to compute the LPC coefficients) should be long enough to detect periodicity, and short enough not to be influenced by far past evolutions.
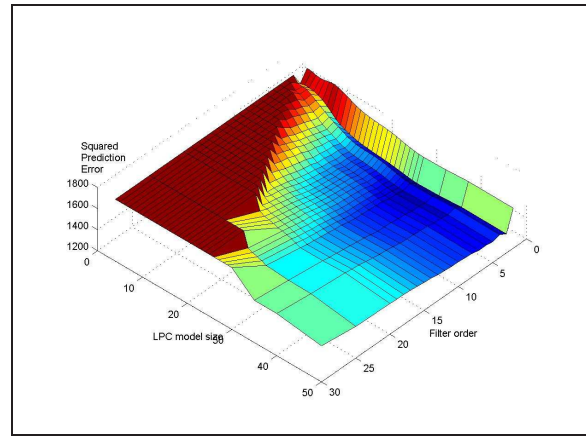


**Figure 4**. Squared prediction error on the next 100 ms, depending on LPC model size and filter order.

The shortest period found in stationary gesture like rolls is around 200 ms. Though, the percussionist play is (hopefully!) far from being stationary, and an empiric study has been performed on a real corpus, trying to minimize for various filter orders and model sizes, the squared prediction error when computing the next 100 ms of signal. We can indeed assume that the body inertia make the movement predictable on the following tenth of a second.

The best results were found for a model size of 30 samples (300 ms), and a filter order of 3 (cf. Figure 4). Although, with such a small model order, the LPC takes only a very small computation time and the real-time implementation is thus straightforward. Since we do not know the percussionist intention, and how correct are the received data, the LPC correction module renders a weighted sum of the raw incoming signal and the softened LPC-synthesized signal. If speed or acceleration increase beyond human capability thresholds, we know that the incoming data are erroneous. We thus extrapolate new samples based on the past, damping the synthesized signal progressively to the mean position, and mixing it progressively with the incoming signal, as it comes back to plausible values.

## 5. GESTURE RECOGNITION

### 5.1. Parameter Analysis

The strikes recognition is based on a statistical analysis of speed and acceleration parameters. After recording a play containing the various types of strikes at various nuances and tempi, we segmented these recordings, and computed the mean and variance of the tangential speed and acceleration on 40 samples (400 ms) window around impacts (cf. Figure 5).

We could reach a satisfying discrimination of the strike types (cf. Figure 6), with the first correlation coefficient between incoming speed signal and the mean we got from the recordings of speed for the various types of strikes.

Though the acceleration would certainly give better results, as it is closer to "intention" than speed, the poor quality and low sampling frequency of the FOB signals prevented us from doing that, because of the Shannon-Nyquist sampling criterion and the twice differentiation
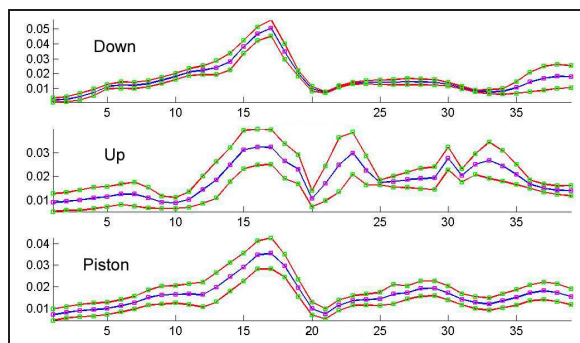
**Figure 5**. Speed mean (blue) and variance (red) around moment of impact for various strikes.

of the position signal which leads to poor numerical precision.

## 6. CONCLUSION AND FUTURE WORK

In the context of an "air percussion" electronic musical instrument, we presented here a data correction and prediction algorithm based on LPC method. Its implementation enhanced the quality of impact detection that is the synchronicity between gesture and sound synthesis, which is fundamental for a percussion instrument. Though it is still at an early stage, statistical analysis of the percussionist strikes in a performance context were realized with shy but promising results, especially when considering the low sampling frequency of FOB and the perturbations.

The use of higher quality sensor system should bring out better results. The idea is to add an acceleration sensor on each stick to obtain a more accurate measurement (higher sampling frequency) of this parameter and combine both the information of position and acceleration. Work still need to be achieved to get a satisfying recognition of the gesture primitives, but once this will be done, it will allow interesting studies the rhythmic patterns emerging in improvisation, with the help of symbolic computation and learning tools. Anticipation and extrapolation could be applied at a symbolic level, and the way would be open for the percussion instrument to all forms of interactions with computers already experimented with MIDI instruments.

## 7. REFERENCES

[1] Havel Ch. and Desainte-Catherine M. "Modeling an Air Percussion for Composition and Performance", *Proceedings of the NIME Conference*, Hamamatsu, Japan, 2004.

[2] Lagrange S., Marchand S., Raspaud, M., and Rault, J-B. "Enhanced Partial Tracking Using Linear Prediction", *Proceedings of the DAFx Conference*, Verona, Italy, 2000.

[3] Kaupinnen J., Kaupinnen I., and Saarinen P. "A Method for Long Extrapolation of Audio Signals", *Journal of Audio Engineering Society*, 49(12), 2001.

[4] Keiler F., Arfib D., and Zölzer U. "Efficient Linear Prediction for Digital Audio Effects",

**Figure 6**. A sample of play with position information (top) and similarity to each strike type (bottom).

*Proceedings of the DAFx Conference*, London, UK, 2003.

[5] Wanderley M. "Interaction Musicien-Instrument : application au contrôle de la synthèse sonore", *PhD thesis* (in French), Université Pierre et Marie-Curie, France, 2001.

[6] Kessous L. and Arfib D. "Bimanuality in Alternate Musical Instruments", *Proceedings of the NIME Conference*, Montreal, Canada, 2003.

[7] Goudard V. "Percussion Aérienne : amélioration de la détection des coups dans le jeu d'un percussionniste", *Master Thesis* (in French), Université de la Méditerranée, France, 2004.

[8] Kippen J. and Bel B. "The Identification and Modelling of a Percussion Language, and the Emergence of Musical Concepts in a Machine-Learning Experimental Set-Up", *Computers and the Humanities*, 2004.

[9] Aslan O. "Le corps en jeu" (collectif d'auteurs), *C.N.R.S.*, 1993.

[10] Wright M. and Freed A. "Open Sound Control: A New Protocol for Communicating with Sound Synthesizers", *Proceedings of the ICMC*, Thessaloniki, Greece, 1997.

[11] Verron Ch. "Captation gestuelle pour une percussion augmentée", *Master thesis* (in French), Université Paris 8, France, 2004.

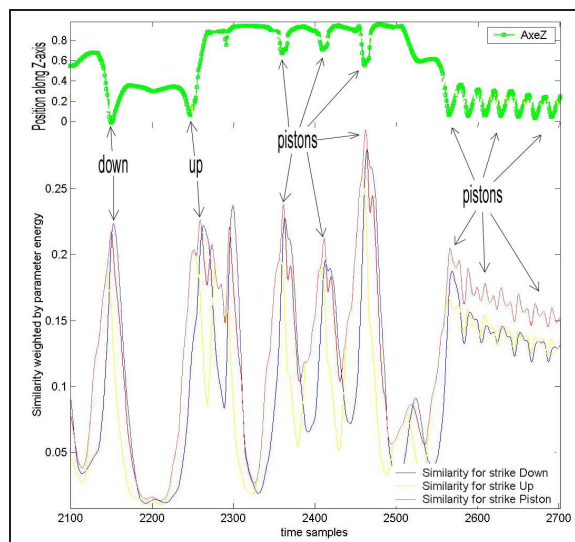[12] Boie R., Matthews M., and Schloss A. "The Radio Drum as a Synthesizer Controler", *Proceedings of the ICMC*, San Francisco, USA, 1989.