Segmentation of Bowstrokes Using the Hidden Markov Model

Aymeric DEVERGIE
Mémoire de Stage de Master 2 Recherche ICA AST année 2004 - 2005
Institut Nat. Polytech., Grenoble

Laboratoire d'accueil: IRCAM, Paris

Equipe: Applications Temps Réel
Responsable: Frédéric Bévilacqua

March / June 2005

Abstract

Augmented violin gives access to an several of instrumental gesture's parameters. We are interested in accelerations of the bow. Those signals contain the major part of informations concerning the instrumental gesture. In particular, we studied the acceleration in the bow axis. Our final goal is to segment bow-strokes during the player's performance. Here, we are using the *Hidden Markov Model* (HMM) in order to recognize and segment different bowing style, as follow: *Détache*, *Spiccato* and *Martelé*. We exposed results we obtained and we discuss possible improvements we may bring to this method to adapt it for more complex experimental data.

Résumé

Le violon augmenté donne accès à de nombreux paramètres du jeu instrumental. Nous nous sommes intéressés aux accélérations de l'archet. Ces signaux contiennent une grande partie des informations du geste instrumental. Nous avons, en particulier, étudier l'acceleration selon l'axe défini par l'archet. Notre motivation finale est de segmenter les coups d'archet pendant la performance du violoniste. Nous utilisons ici les Modèles de Markov Cachés (HMM) afin de reconnaître et segmenter différents type de jeu, à savoir: *Détaché*, *Spiccato* et *Martelé*. Nous exposons ici les résultats obtenus et les améliorations futures à apporter à cette méthode pour l'adapter aux données expérimentales plus complexes.

Acknowledgments

I will acknowledge Norbert Schnell for his welcome in the 'Real Time Application' team of research.

I would thank my supervisor, Frédéric Bévilacqua who proposed to me this interresting topic of research, and who gave me lot of precious advise.

My collabaration with Nicolas Rasamimanana was fruitfull. His motivation was very communicative. For his support, I will thank him. Discussions we have were so interresting.

At last, I will acknowledge all collegues who support me everyday: Frédéric Leau, Julien Rachedi, Julien Bloit, Philippe Zemar, Baptiste De la Gorce, Patrice Tisserand and Isabel Pires.

Contents

List of Figures

Chapter 1

Introduction

Since twenty years interest for gesture increased consequently. For years, the only aspect studied in instrumental research was the sound. Naturally, it seemed to be the most important aspect to study, and it is. Analysis and ,later, synthesis focused on sound and acoustical proprieties to create musical material. To control sound synthetized with those signal processing methods, the controllers employeed are the mouse and the keyboard. In MIDI interfaces, the control is obtained with minimalist information. A number represents the beginning, the end and all low level informations.The controllers derived from computer's technology reduce possibilities of an expressive and natural control. Instrumental gesture is gesture performed by any instrumentist when he is interacting with its instrument. In [1] and in [2] purposes are centered on instrumental gestures. Goal of research on gesture are numerous: increase our knowledge on the instrumental gesture, develop new interfaces and integrate gesture in new process of artistic creation.

Violin represents a special case in instruments. Its family of instrument is particular. The sound is produced thanks to a permanent contact between the vibrating structure: strings, and excitating device: the bow. Sounds are, however, particularly connected to gesture proprieties. Plucked instruments are not so sensitive to instrumental gesture. Once the vibrating structure is in movement, influence of the exciting gesture is different from the case of violin.Thus, violin becomes an interesting subject of research for the instrumental gesture.

At IRCAM, gesture acquisition became an important subject of interrest since 2000. Prime works where achieved by Marcelo Wanderley in 2000 and in 2001. His contributions is mentionned in [3], [4]. Marcelo Wanderley related his work on gesture control of music and on performer - instrument interactions. Emilie Morin in [5] and in [6] related his work on analysis of gesture features on violin. Gestures acquisition used the *Digibow* system. This prior work gave concrete solutions for datas acquisitions. Thoses publications gathered all knowledges on the current topic of research.

Since 2003, a new team of research focuses on analysis and technologies in musical and dance performances. Work of Emmanuel Fléty and al. related in [7] in 2004, are developpement's works. Existing system of datas acquisition where limited. Hardware developpements of sensors in term of performance, size and weight were achieved. Thanks to this technological improvements, datas revealed itself more precise and more significative.

In the context of the gesture segmentation in music and dance performances, our work consist in finding a method to segment different gestures of the violonist's play. Work achieved by N.Rasamimanana and al. [8] is the starting point of our reflexion. This previous work revealed that bowing gesture of the violinist could be identified with analysis based on accelerations of the bow. Three different bowing styles were studied in this master report: the *Spiccato* , *Detaché* and *Martele* bowing styles. The bowing techniques where characterized in a off- line process. Gestures vary from a player to another, from a tempo to another and depends on the acoustic level of the performance.

They found significant parameters to characterize bowing styles on datas manually segmented. To continue in this direction, we decided to work on a segmentation method of the bowstrokes. The main difficulty for a segmentation of the strokes, is to find the most significant parameters in bowing gesture necessary to characterize them. Parameters must be enough significant and not too consequent to be analysed in real time. Considering a single cartesian direction of bowing gesture, the playing direction, we use in this report a statistic method, called the Hidden Markov Model.

Parent domains of research are linked to our problem of data segmentation. We here mention work made in speech recognition and in Electrocardiogram analysis. In speech recognition, the mainly used *segment* is the phoneme. Decomposition of phrases in phonemes in similar to a segmentation of bow gestures in violin play. Phonemes contain articulations of the speech, bow stroke contain articulations of violin game. A second topic is the ECG analysis. The segment are precise patterns of heart pulsations. In this domain, pattern are similar to bow strokes but ECG pattern are far more periodical than bow strokes. The datas lightly differs between two different persons. This is not the case for gesture datas. ECG analysis, give us lot of ideas for our problem of segmentation.

The aim of segmentation of instrumental gesture applied to violin may be the score following. Nowadays, score following is based on acoustic features. In future, gestures identification may be used to follow the score. Gesture proprieties may be used to explain sounds properties. Spectral informations can be linked to gestures proprieties.
Segmented gesture may control video or other acoustic materials, in real time performances, for example.

First of all, we sum up in section 2 the litterature concerning our topic of research. In section 3, we introduce datas we are going to study. In section 4 and in section 5 we present our experiments results. In section 6, performances of the method are presented. And at last, in section 7, we suggest perspectives and future work.

Chapter 2

State of The Art

In the following section, we discuss topics directly linked to our problem: the segmentation of bowstrokes on an augmented violin. Signals processing is described in highlines. Then we introduced the Acoustic violin research's domain.The question of data segmentation is an specific domain, where lot of applications already exist. At last we present some multimedia installations combining all previous domains we present.

2.1 Computer based Analysis

Two different approaches exist to analyse sounds or gesture. To analyse sound and every signal we record datas and analyse it after.This approach is the off-line analysis. The second method is called real time analysis. In order to analyse signal, it is recorded and treated at the same time.

Before describing the two approaches, we will discuss about the improvements in the domain of computing. One essential tool for analysing signal is the computer. Analysing a signal consists in making a lots of math operations. At the beginning of computer's era, processors were very slow and memories were small. In this context, nobody would pretend to do real time analysis.

Fourier proposed theorems to describe the signals. These theorems were demonstrated before the outset of computing. But thanks to computers, we were able to analyse the signal with spectral considerations brought by Fourier's theorems. Lost of analysis are based on Fourier' s methods. The spectral approch is fruitful to understand acoustic aspects of signals. Fourier's analyse is one of the most used methods in off-line analysis.

The computing ressources allows to do complex operations, and solve difficult problems. The spectral analyse lead by Fourier's method may be realized with one single personnal computer. Fourier's decomposition brings the spectral distribution of every periodical signal. Adaptated methods work for pseudo and non periodical signals.

Two different methods are frequently used to analyse non periodical signals.The first one is called the Principal Components Analysis. It is used to reduce the dimensions of a dataflow. If we want to show the most important characteristics of a phenomenon, we use it. The PCA method reduce datas in order to find the most important directions where the datas are changing. Directions are the principal ones where we observe the bigger variance of datas. A second approach, similar to the PCA, is the Linear Discriminant Analysis. This method differs in the finding of directions. The directions found are the ones where we can discrimate the most the datas.

To perform real time analysis of physical signals, methods should be efficient. They are supposed not to need lot of computing ressources. Methods are mainly statistic methods. In [9], B. Schoner, syntetize sounds from physical datas extracted with a probabilist method. He achieves to classify entry datas in real time. Gesture of the violionist are treated and compared to datas classes. Each type of datas are gathered in clusters. Q comparison is made between incoming datas and types of gesture registered before. Once the comparison is made synthetic sounds are produced. Degrees of freedom of the violonist are limited, but allow to control virtual enviromment in real time. Schoner' works inspirate itself from a well-known method: *Hidden Markov Model*. We are refering to [10] for further explanations.This is a mixed method between statistic and learning methods. This methods extract from datas differents states in which the system can be. System can go from a state to another with a given probability. The state is not clear, it is itself described by a probabilistic distribution. Once we have declared the states and all the probabilities, the system can be learned. We show it different succesions of states, and it modify the transitions probabilities in order to describe in the best way the experienced evolution of the system. After the learning stage, it is capable to recognize different states of the system. Such a method is very commun in the domain of voice synthesis and speech recognition.

2.2 Acoustic Violin Studies

In the domain of acoustic research, the violin is a particular case. Complexity of sounds produced and gesture possibilities constitute a large interesting aspect of acoustics, physics and signal analysing. The permanent contact between the bow and the strings is creating complex relationship between the player and its instrument. To understand how the violin produce sounds, we can study the real violin in different ways.

Spectrum analysis is very interesting in this case. The tone color of sounds can be revealed when we examinate the spectrum. The tone color is a consequence of the spectral content of the signal, as an illustration, see in [11], works of K.guettler. Transients are the most important proprieties of sounds, which let us recognize the instruments. Suppressing initial transients and we will not identify a violin when we are listening to a blow instrument, for example.

Violin is producing sound according to the stick slip motion. During a fraction of the vibrating period, the bow is sticking to the string thanks to rosin. During the rest of the period, the string is moving free. Such a succesion of stick slip periods is repeating until the bow stops. This non linear interaction is specific to the violin. We can explain how the playing parameters change sounds. Most of the studies are still based on spectral analysis. In the violin playing, the bow velocity, the bow position and the bow force are the most influencing parameters. Askenfeld studies those parameters in [12]. Bow force applied on the strings influences the moment when the string slips under the bow. The stronger the force is, the later will this moment appear. Askenfelt discovers that the first most influencing parameter is the bow position along the string. This is different from the transversal position of the bow on strings. A short distance from the bridge, produces a tiny and slight sound. The string can not move widely. In the other typical case, the string is played far from the bridge, so that it can vibrate with the maximum amplitude. It sounds raucous and low modes are more present. The bow velocity changes the tone color. The faster you play, the lighter it sounds. But during a slow motion of the bow, the tone is sharp and brilliant. The bow force influences the spectrum as the position can do but more lightly. When the bow force is high, the tone is raucous, in the other case it sounds brilliant.

2.3 Augmented Violin

In the research on violin, some experiments will expose limits of investigations we can make with real instruments. In order to find more details on the violin charactiristics we expend the violin by placing sensors on it. Augment signals with sensor is a part of the augmented reatlity. We observe reality of the violin with devices able to reveal informations we could not 'see' with our sensitivity. The augmented reality is the next step where datas are used to control or interact between human and computing domains. In [13], authors draw main directions of evolution from interaction with physical objects to augmented reality.

As an introduction a *virtual violin* may be considered as a collection of datas derived from physical process of playing the violin. The first virtual violin is the 'Envelop tracker'. The envelop of a real electric violin signal is used to modulate another signal. In 1970, Laurie Anderson replaced the horse hair of the bow by a magnetic tape. The violin was fitted with audio tape recorder heads. The prerecorded material on the tape was sent to an amplifier, and was modified by the play of the violonist.

The pitch detection consists in quantifying the only audio signal of a source. It is not considering physical processes which produce signals. M.Pucket developped a particularly efficient method for sustained pitches and multiple pitches simultaneous detection [14].
To conclude, such methods of envelop and pitch detection consider the violin as a mechanism of control. It is coordinating conventional violin performance with precomposed pieces.

In [15], Tod Machover, Neil Gerschenfeld and Joseph Paradiso are presenting a different approach. They have built the *Hyper Instrument* at the MIT Labs. They built variety of sensors and mounted them on instrument to measure performance gestures. The Hypercello consists in a sensor bow. This bow measure the finger pressure, wrist and bow position, and otherwise left hand fingerboard position. The hypercello send its datas through a network of computers for analysis and control of synthesis. Diana Young developped a Hyper controller. Her works are related in [16],[17] and [18]. The acceleration sensors are placed at the frog to. Position sensors are made of a resistive strip attached to the stick of the bow. Downward and lateral strain are measured. The system transmit datas with a wireless system.

In 1980, at STEIM Jon Rose built a Hyper Violin with pressure sensor to measure finger pressure on the bow, [14]. A sonar sensor was developped to detect bow position. Later, Chris Chafe, augmented a cello bow, [14]. He placed bend sensor in the inside of the bow stick, and accelerometers at the frog. C.Chafe uses two force sensing resistors placed between the stick and hairs of the bow. At the frog, he puts a bidimensionnal acceleration sensor. The 'BoSSA' was designed by Perry Cook, [14]. It differs from the system mentionned just before in the bend sensing.

2.4 Datas Segmentation and Analysis

In general, we are interested in data segmentation in order to analyse signals. And in a second time, segmented signals could control synthesis techniques or other kind of computer based processes.

The first domain of data segmentation is the domain of sound signals. One of the most used method is the 'Onsets, Offsets Detection'. The aim of this method is to caracterize musical objets. One objet is composed from an attack, a decay, a sustain and a release. In [19], the author discuss his real time detection of sets of musical objets. He tackels the problem of real time low latency requirements. This object based construction allows to build spectral model of musical instruments. Moreover, the onset detection is limited for gesture datas. For a same gesture, its magnitude variance and time warping brings the detection of onsets difficult.

As a reference, we mention the chapter entitled *'Gesture Based Interaction'*, in [20]. In this chapter, Mark Dillinghurst presents an epistemology of the human gesture. In a second part he lists the different types of the Gesture based interfaces. In order to perform a gesture segmentation different methods are available. A epistemology of gestures is presented in [21].

Some works were made to segment gestures. The work of Chad Peiper and al. [22] exposes a computing method to classify and to segment gestures. The program routine classify the gesture inputs from a violonist. The algorithm is based on a decision tree witch analyse bow parameters and stores them in different classes. The decision tree is evoluting thanks to a learning method. In spite of the existence of the learning, some case are difficult to classify. The system works efficiently for standard gestures. Once a singular gesture is achieved, the system could not put it in a classe. A parent method is proposed by Bernd Schoner and al. in [9]. This method considers clusters of bowing parameters. Each cluster is representing a kind of bow stroke. This method may solve problems encountered by the previous method. But it is still reducing the variety of bow strokes by putting bounderies between kinds. After work of clustering is made, Bern Schoner's method controls waveform sound synthesis.

In [23], it is exposed a method for analysing expressive gesture in dance and musical performance. It is related that the movement could by analysed at different levels. This question arise in our context too. Low level may be for a dance performance the movement of the hand for example. The high level of the segmentation may be the intention in the movement.

Closer to our objectives, two researches are interesting. In [24], muscle's hand tension of the instrumentist is used as a cue of segmentation. Variations of electromyogram signal is used to segment gesture. In [25], accelerations sensors are employeed to analyse gesture motions.

2.5 Gesture interface in virtual environnements

Before starting, we may refer to the article of Kristoffer Jensen [26] in which he lists all the possibilities we have to control sounds when we play musical instruments. We can give some examples: loudness, pitch, tone color, noise and inharmonicity. It is interesting to know how we can modify sounds in order to develop virtual interfaces.

The first step achieved to control virtual environnements is the data-driven model. First, datas from real instruments are gathered, then they are analysed and finally, they are used as input parameters to control a virtual instrument. In [27], the virtual instrument is a bowed string synthetized by a wave guide model. The input parameters are the bow position , the bow velocity and the bow force. Those paramaters control a computer animation of a virtual violin player. In [28], another approach is presented. Parameters are cluster weighted and store in different classes. Then, they control the synthesis of acoustical instruments. Some new parameters are taken into account: the pressure of the player's finger on the bow and the acoustic input of the violin.

Those interfaces are not considering the gesture as a parameter. In [29], Sylviane Sapir exposes a new kind of research in computing technology. The interest given in human gesture interface is increasing since real time device are available. Developpement of tracking system introduce the gesture as a way to control virtual instruments. Gesture can be considered as a new musical paramater to play with in a context of musical performance. S. Sapir presents the technology of a gesture tracking system in details in his paper . An application of such a kind of systems, developped by Max Matthews, is presented in [30]. The application is called 'The Radio Drum'. This is a real time device where input datas are the gesture of the player. Sticks in his hands are placed in front of radio frequency sensors witch are detecting variation of the field produced by the player's movements. Gesture may be capture in a different way. Lot of devices are video based. In [31], such a system is described.

A new domain of research witch differs from what we presented just above is the domain of 'The new interfaces'. J.M.Couturier [32], presents is interface witch controls a Bow String model. It consists in a tablet, a touching sensor surface and a screen. Another interface is presented in [33]. The interface described is similar to a real violin. It is called ' The SuperPolm' . There is no strings mounted on it. It is made of different kind of sensors.

In 1995 Curt Bahn and Dan Truman performed a musical piece in witch they played interfaces similar to the SuperPolm. This experience is related in [34]. C.Bahn played a augmented bow and D. Truman played a Bass with sensors. This experience was one of the first performance made in this context.

Chapter 3

General Description of the Signal

3.1 Method of work

The displacement of the bow in the playing direction is studied here. Both position and acceleration of the bow are collected thanks to sensors placed on the bow. Our research work concerns the acceleration datas of the bow. Those datas characterize the instrumental gesture. Dynamics of gesture are more interesting since sounds produced by violin are direcly influenced by bow velocity. Direction of acceleration, considered, is the playing direction, perpendicular to strings and horizontal. We study three different bowing style: *Détaché*, *Spiccato* and *Martélé*.

Datas are sent through the *Ethersense system* to *MAX/MSP*. This point is developped in chapter A page i.Then datas are converted in text files. Those text files are loaded in *MATLAB*.

We work with MATLAB Software. We develop scripts in this langage for a theoric validation of ideas for a segmentation of the bowstrokes. In the same time, ideas revealed interesting by our work are computed in *MAX/MSP*. This work is made in order to make real time segmentation of strokes.

3.2 Definition of a Bowstroke

Acceleration of the bow is plotted time-magnitude graphics. In figures 3.1, 3.2 and 3.3 three different bowing styles are represented. X axis corresponds to the sample's number. The sampling frequency is equal to 250 Hz. Y axis is a Arbitrary Unit.

3.2.1 *Détaché* strokes

Figure 3.1 shows a succesion of a *Détaché* Upstroke and a Downstroke.

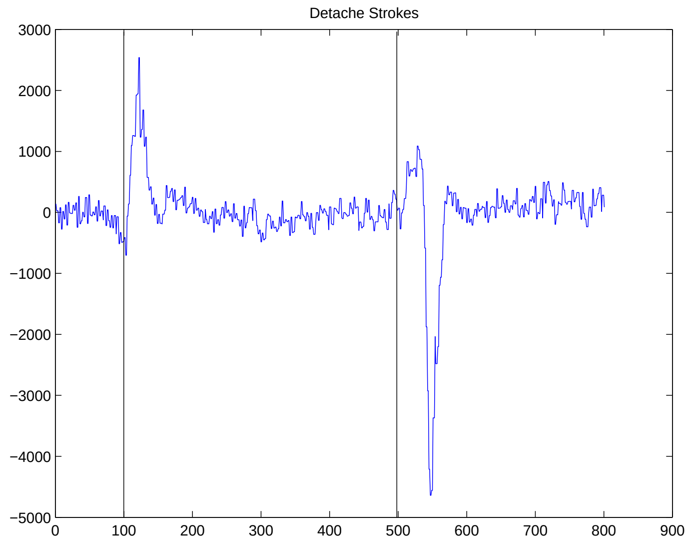


Figure 3.1: *Détaché* strokes.

Détaché stroke is the simpliest stroke. When the acceleration rises, it is an Upstroke. When it drops, it is a Downstroke. The nearly 'constant' part of the signal corresponds to a constant velocity of the bow. Variations are due to bow-string interactions.

Plot reveals an anticipation gesture. Before starting the downstroke, acceleration lightly increases. The hand of the violonist gives the impulsion to start the gesture.

3.2.2 *Spiccato* strokes

Figure 3.2 shows a succesion of *Spiccato* Upstrokes and Downstrokes.

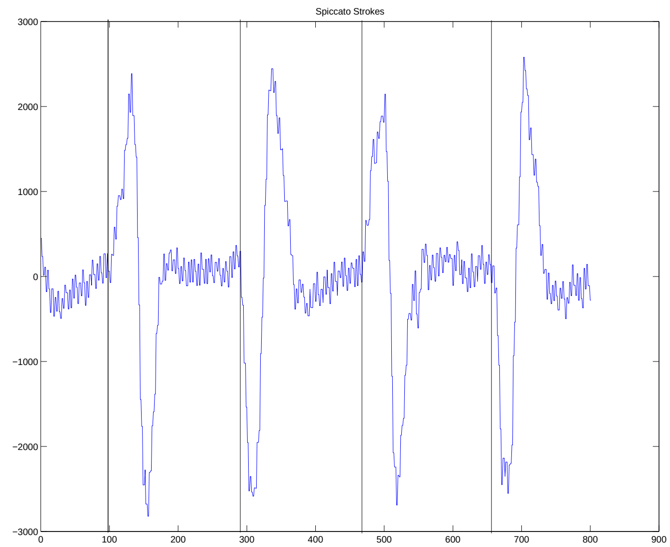


Figure 3.2: *Spiccato* strokes

Spiccato is a constrained gesture. Acceleration and Deccelerarion are similar. This brings a symetric aspect to the curve. The initial direction of acceleleration curve informs us if the stroke an Up or a Downstroke is.

3.2.3 *Martelé* strokes

Figure 3.3 shows a succesion of *Martelé* Upstrokes and Downstrokes.

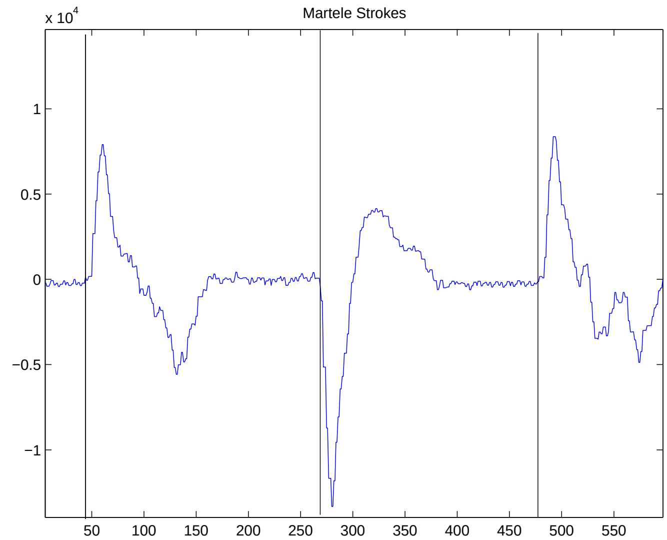


Figure 3.3: *Martelé* strokes

Martelé stroke, is non symetric. The dominant direction of the gesture is the expected direction. A residual decceleration appeared to stop the bow at the end of the stroke.

3.3 Analysis of likelihood curve

3.3.1 Protocol

The *Spiccato* Downstroke, in blue in figure 3.4 is the shape tested. The aim of the experiment is to qualify the evolution of the likelihood curve when HMM is supposed to recognize the shape of the acceleration curve.

3.3.2 Description of the figure

The reference is a *Spiccato* Downstroke, in red in figure 3.4. The likelihood is evaluate between the blue and the red plot. The black curve represents the log probability as mentionned in section B.2 on page vi. A substantial information is contained in the differential of this likelihood curve. It is represented with the green plot.

The likelihood plot represents the log probability. The closer to 0 it is, the better the pattern is matching to the reference. If likelihood drops, it means that probability falls. The HMM no longer fit with the tested curve.

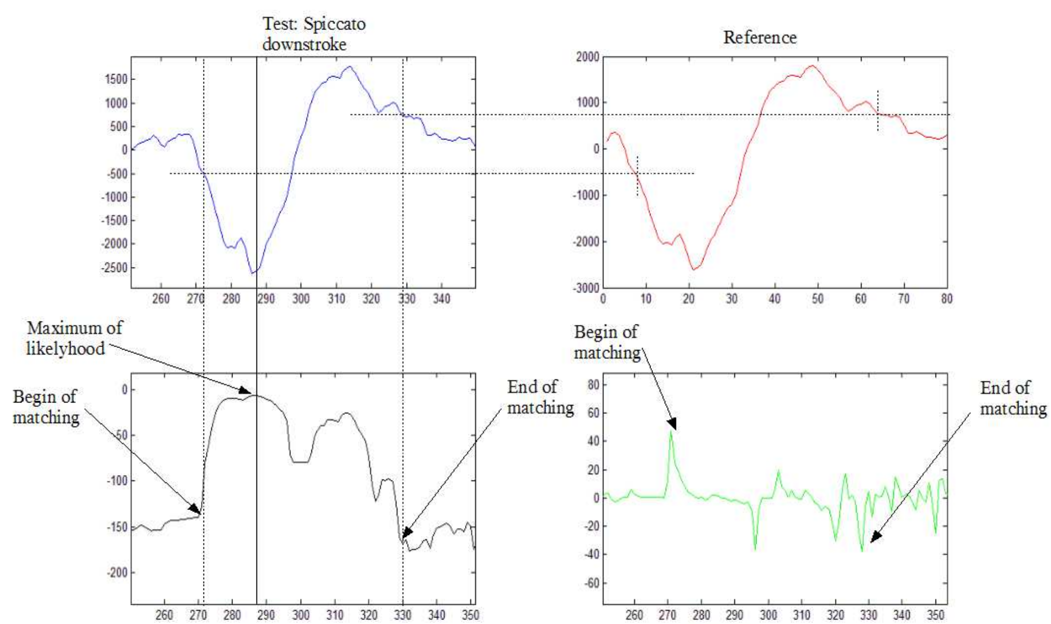


Figure 3.4: Explanations of the likelihood curve

3.3.3 Interpreting the likelihood plot

1. Evolution of the likelihood:

   · When the HMM is not fitting with the test, the likelihood remains low .

   · Once the beginning of the Test matches with the HMM, the likelihood curve increases rapidly.

   · Maximum value of the likelihood curve is reached when Test and HMM match perfectly.

   · Once HMM is moved out of the 'matching window' the likelihood curve drops rapidly.

2. HMM is similar to the test when:

   · Likelihood curve reaches a maximum.

   · When likelihood rises and drops rapidly with in between a 'stable' decreasing period.

   · In the differential of the likelihood curve, a specific pattern appears. A short durating succession of a positive spike, constant period and a negative spike, reveals the same information as item mentionned before.

   · Considerations made on the differential curve are sensible.

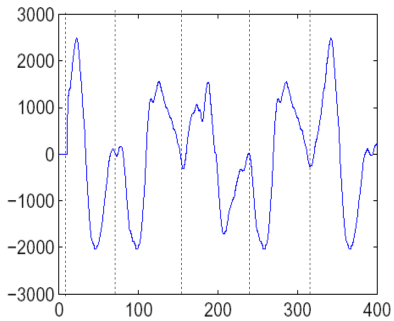
Chapter 4

Stroke recognition using a Standard HMM

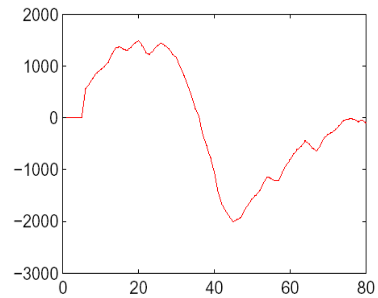
The HMM we name *Standard HMM* is described in section B.2 page vi. The *Emission matrix* is described with a constant standard deviation for the gaussian distribution.

4.1 Tested sequence and Reference

Datas are filtered with a Savitsky Golay filter [35].The reference is a *Spiccato* Downstroke. The sequence analysed with the reference is a succesion of *Spiccato* Upstroke and *Spiccato* Downstroke.



(a) Test sequence: Spiccato Strokes



(b) Reference: Spiccato Upstroke

Figure 4.1: Standard HMM: Test sequence and Reference.

In figure 4.1(a) dashed line separates an Up stroke from a Downstroke and so on. The first pattern is a *Spiccato* Upstroke. In figure 4.1(b), the reference is a *Spiccato* Upstroke.

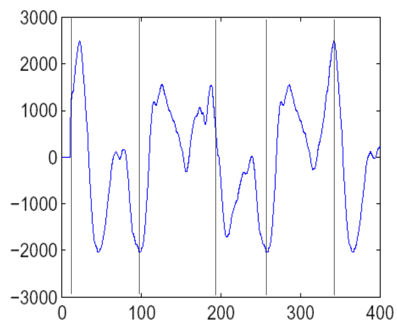
Note that the figures 4.1(a) and 4.1(b) are plotted with different time scale, but the duration of th reference and the test strokes are on the same order.

4.2 Likelihood Evaluation

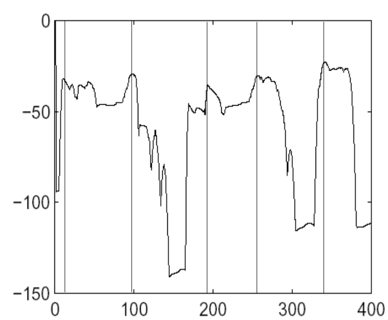
The value of likelihood at a time t corresponds to the instant where the first sample of the reference is matching with the first sample of the window of observation. Window of observation is a part of the tested sequence witch duration is equal to the reference. The reference is moved sample by sample from the beginning of the tested sequence to the end of it.

4.2.1 Likelihood

Figures 4.2(a) shows the sequence of Bowstrokes we test and the Likelihood 4.2(b)



(a) Sequence 'segmented' thanks to likelihood



(b) Likelihood

Figure 4.2: Likelihood for Standard HMM

Analysing results for Likelihood

A vertical stem is placed where the likelihood is locally maximum. Sequence of stems are reported on the sequence.

A vertical stem corresponds to the instant when the reference is supposed to match with the test. In figure 4.2(b) the second and the fourth stem are not corresponding to the expected recognition. HMM makes mistakes.

Positive part of acceleration curve for the Upstroke similar to reference is recognized. Negative part of the curve for the Downstroke similar to reference is recognized. No distinction is made between the stroke supposed to be recognized and its opposite (not supposed to be recognized) . Likelihood is potentially as good for a similar stroke as a different stroke. The recognition of the HMM is not satisfying.

Moreover, a lighlty different Upstroke (for a reference: Upstroke) changes the estimated likelihood. The value of likelihood is smaller. A constant standard deviation, equal to 200 a.u, for gaussian distribution is the reason why we observ such differencies between same strokes.

4.3 Scale problem and suggestions

The scale is the main problem. A constant standard deviation for the *Emission Matrix* brings the HMM not enough flexible to scale's variations between Test and Reference. For a given state of the tested curve, magnitude variations brings HMM not recognize its corresponding state in the referenced sequence. Two opposite directions to solve this problem are exposed below.

We achieved some interesting methods to rescale the analysed datas. The first idea constists in rescaling the signal between 0 and 1. Such a method suppress information contained in the ratio between min and max of the curve. Second idea was to place the mean value at 0. The minimum is set to -1, and the maximum remained free. If we did so, a sequence with different bowing styles became problematic. The magnitude differencies between two different bowing style, make the HMM not efficient in the recognition process. After discussing rescaling methods revealed itselves limited. We can easily find a case where our rescale's method will not work.

Opposite consideration is the modification of the HMM. Instead of finding a method to fit datas to the model, we fit the model to datas. In the following, we introduce Autoregressive Model.

Chapter 5

Autoregressive HMMs

A standard HMM is non causal. In a sequence of states, the next state is not depending on past states. Each state is independant.

With autoregressive model, states are linked. In physical process like Bowing gesture, acceleration is a continuous phenomenon. The value of acceleration at a time t depend on the value at a time t-1. Autoregressive model considers the state at time t and the influence of the former states. The past of the signal is a additional a primal information used by the modelisation.

5.1 Autoregressive Model

Principe of Autoregressive Model (AR) is to estimate the sample at a time t, knowing the past samples. To do the prediction, a linear predictive method is used.

5.1.1 Yule Walker Algorithm

We used the Yule Walker method ([36]) at the first order. The order signifies how many past samples we take in account to estimate the new sample. The YW algorithme evaluates coefficients, named the AR coefficients, wich let us evaluate the next step sample.

The estimated state's value is:

$$\hat{S}_j = \Phi_1 \cdot S_{j-1} + \xi_{i+1} \tag{5.1}$$

Φ_1 : AR coef., S_j : State of the HMM, ξ : Prediction error variance

We see in figures (5.1), (5.2) below an illustration of estimation process.

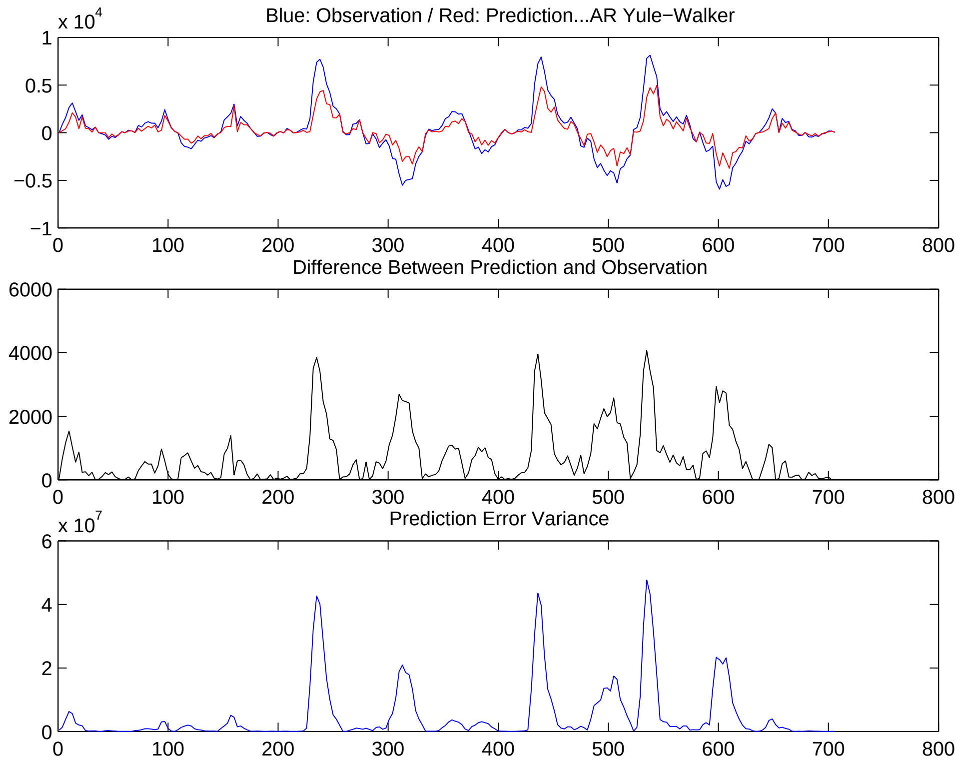


Figure 5.1: Yule Walker Estimation method, order 1

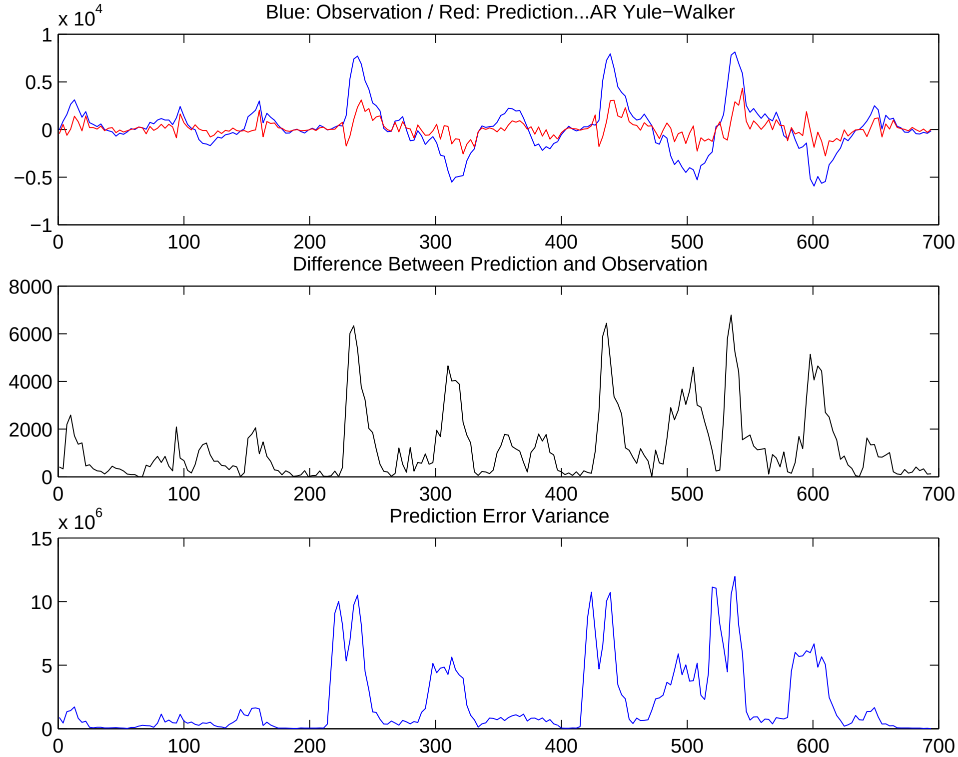


Figure 5.2: Yule Walker Estimation method, order 5

The more adapted order to shapes we study is the first order. Prediction error variance is worth for a superior order YW algorithm.

5.1.2 Properties of the AR Model

Autoregressive model is introduced to redefine the gaussian distribution in the *Emission matrix*. The mean values of gaussains are no longer mean value between two successive states, but become the difference between the state at time t, and the estimation of it, made at time t-1. The reference no longer describes the acceleration curve, but a difference curve between estimation and observation. The fundamental information is the evolution and not absolute values of the curve. This brings HMM robust to magnitude variations of curves.

The standard deviation of gaussians is the variance of prediction's error. When the observed state differs from the estimated state, the prediction error is important. When it is close to estimation, error is small. In figure 5.3 , two different strokes are plotted. Differences between them are only a magnitude difference.

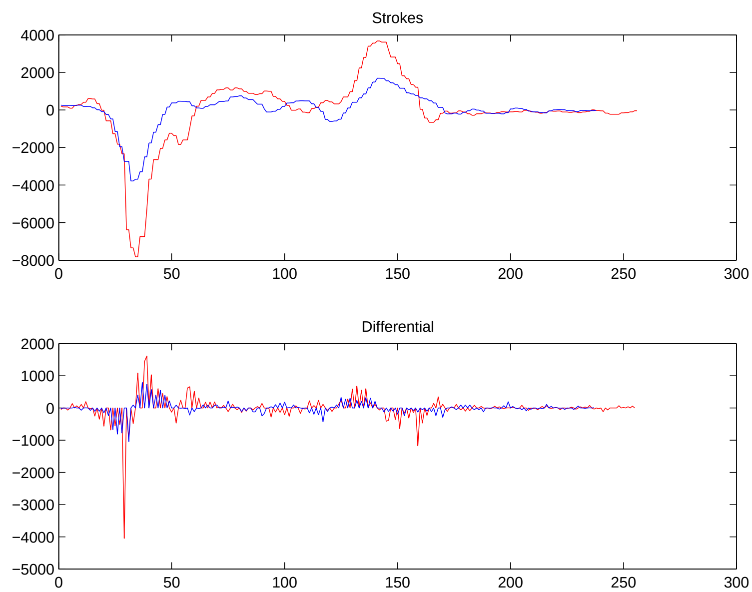


Figure 5.3: Magnitude differencies between a mezzoforte and a pianissimo Bowstroke

At extremum values of accelerations, differencies are significant. With AR model, the prediction error variance is important. The prediction error variance is small elsewhere. The model is flexible at extremum values, and constraint otherwise. It is perfectly adapted to this situation.

5.2 Modification of HMM caracteristics

Parameter we modify are the $b_j(k)$ coefficients (section (B.2), vi). Those coefficients are the *Emission matrix* coefficients. likelihood evaluation is linked to them. In [37] the refered method is explained in more details.

Here, we consider the past observation to affect the *Emission* coefficients. It will straight the most probable state, considering the past state of the frame. The 'less probable states' contributions are reduced.

Secondly, prediction error variance defines the new Sigma of gaussian distributions.

$$b_j(k) = \frac{1}{\sigma' \sqrt{2\pi}} e^{\frac{-(v_k - \hat{S}_j)^2}{2(\sigma')^2}}, \quad 1 \leq k \leq N, \quad 1 \leq j \leq T. \tag{5.2}$$

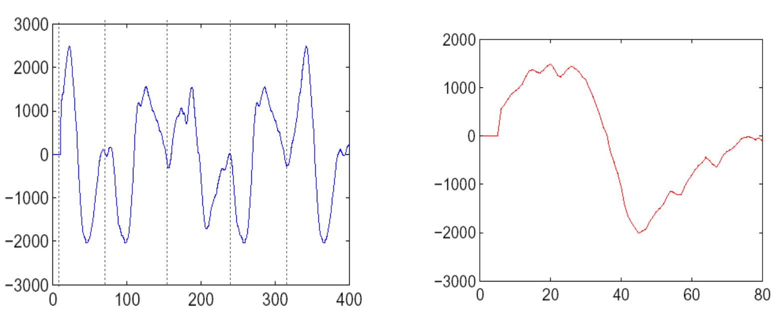
$\hat{S}_j$: estimated state define in [5.1], σ' : $\sqrt{\xi} + cst.$ define in [5.1]

$cst.$ is an additionnal margin set to give HMM flexibility in the recognition process.

5.3 Results with AR Model

5.3.1 Test and Reference

The protocol is the same as in section 4 page 17. The tested sequence is a succession of *Spiccato* strokes. The reference is a *Spiccato* upstroke.

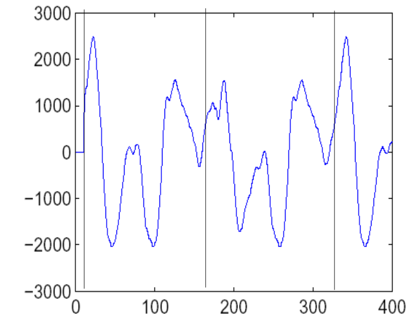


(a) Test sequence: *Spiccato* Strokes (b) Reference: *Spiccato* Upstroke

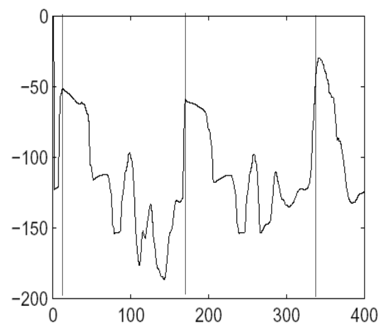
Figure 5.4: Autoregressive HMM: Test sequence and Reference

5.3.2 Likelihood

Figures 5.5(a) shows the sequence of Bowstrokes we test and the Likelihood 5.5(b)



(a) Sequence 'segmented' thanks to likelihood



(b) Likelihood

Figure 5.5: Likelihood for Autoregressive HMM

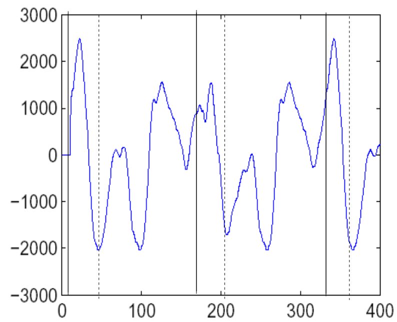
Analysing results for likelihood

The likelihood reaches a maximum when the HMM perfectly matches with the tested sequence. likelihood drops then constantly. Finally it drops rapidly when the HMM no longer fit with the test. This particular shape of the likelihood curve represents an ideal shape.

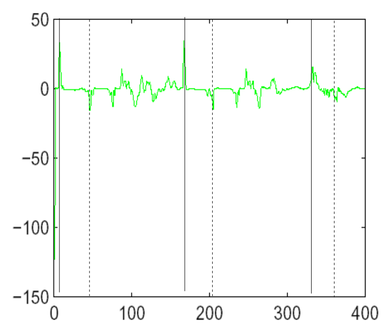
In the sequence we test, there are three occurrences of the reference. Stems we placed sort, in this sequence, good occurrences of the reference. The AR model is efficient with apperently difficult shapes like biphasic *Spiccato* shapes. Compared to the standard HMM, results are better.

5.3.3 Differential of likelihood

Figures show 5.6(a) the sequence we test ,figure 5.6(b) the differential of the Likelihood .



(a) Sequence 'segmented' thanks to Differential of likelihood



(b) Differential of likelihood

Figure 5.6: Differential of likelihood for Autoregressive HMM

Analysing results for Differential of likelihood

The differential is enough precise to extract some informations. The typical sequence of a positive spike, constant period and negative spike, is clearly observable here. Positive spike occurs when the HMM is in phase with the test. The negative spike occurs when the two shapes are in quadrature. The constant period corresponds to the constant decreasing of the likelihood curve.

Plain line underlines phases period. Dashed line underlines quadrature phases. Differential of likelihood should be consider with less weitgh than likelihood. Experiments should be made to conform us in the validity of informations contained in it.

Chapter 6

Estimation of the AR Model performances

In the following section, datas are musical scales played on the A and E string. Strings are played *mezzoforte* in the three different bowing styles: *Détaché*, *Spiccato* and *Martelé*. Small variations of acceleration magnitude introduced with real datas is a new interresting specifity.

6.1 Sequence of Same bowing style

The sequence contains an Upstroke and a Downstroke. For each bowing style, we test the sequence with the stroke of the corresponding bowing style. We focused on likelihood curve.

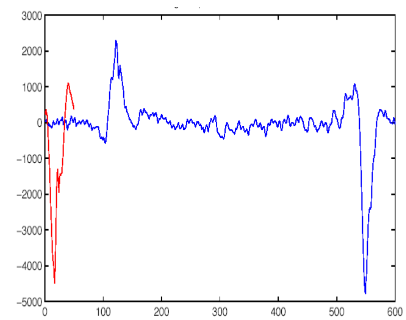
6.1.1 Description

The upper left figure represent in blue the tested sequence and in red the reference we expect to find in the sequence. The upper right, is the tested sequence. The lower left figure is the likelihood between tested sequence and the reference. The lower right figure is a focuse on the most important information contained in likelihood curve. The zoom is made between start and end point of the sequence. Only the Y axis is zoomed.

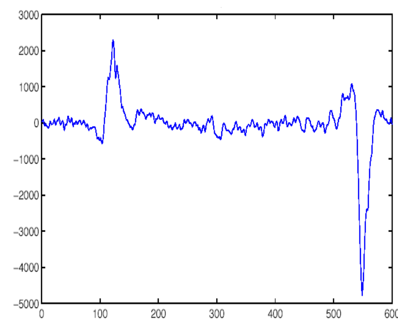
Values outlined in figure, are local maximums of likelihood. The test is matching with a part of the reference or with the whole one regarding the value of likelihood.

6.1.2 Figures: ***Détaché*** strokes

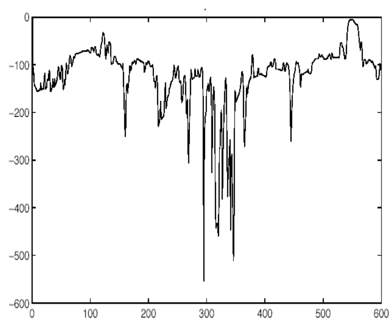
Figures 6.1(a) , 6.1(b) shows the tested sequence. Figure 6.1(c) shows the Likelihood, and figure 6.1(d) a zoom on Likelihood.



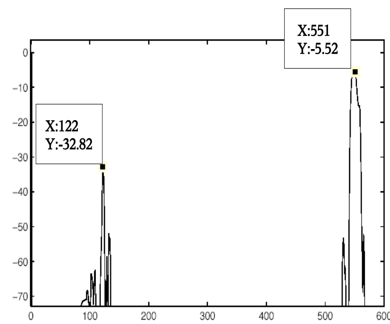
(a) *Detaché* strokes (blue), *Detaché* Downstroke (red)



(b) *Detaché* Strokes



(c) Likelihood



(d) Zoom of likelihood

Figure 6.1: Likelihood for *Detaché* Strokes

6.1.3 Results: ***Detaché*** strokes

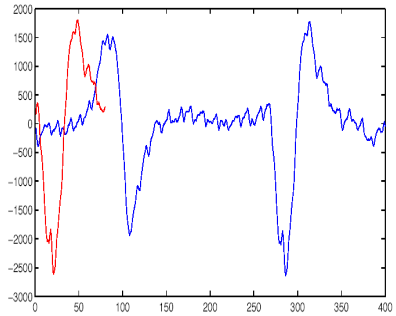
· Likelihood is equal to $y = -32$, for the opposite stroke

· Likelihood is equal to $y = -5$, for the corresponding stroke.

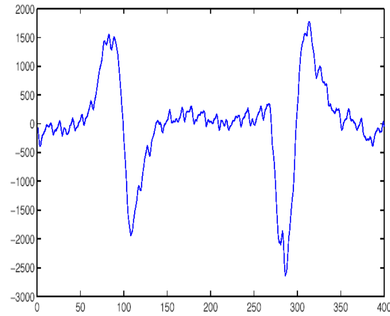
The closer to 0 it is, the better matching it is. For this bowing style, the HMM recognizes the right stroke. Differencies between values are significant.

6.1.4 Figures: *Spiccato* strokes

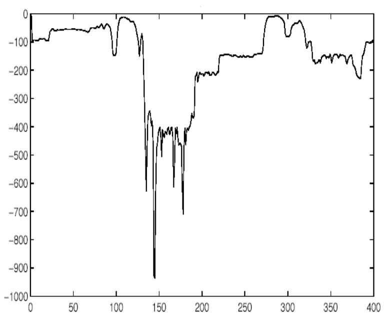
Figures 6.2(a) , 6.2(b) shows the tested sequence. Figure 6.2(c) shows the Likelihood, and figure 6.2(d) a zoom on Likelihood.



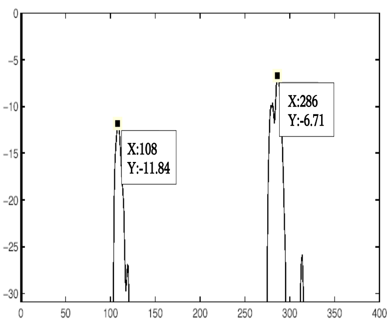
(a) *Spiccato* strokes (blue), *Spiccato* Downstroke (red)



(b) *Spiccato* Strokes



(c) Likelihood



(d) Zoom of likelihood

Figure 6.2: Likelihood for *Spiccato* Strokes

6.1.5 Results: *Spiccato* strokes

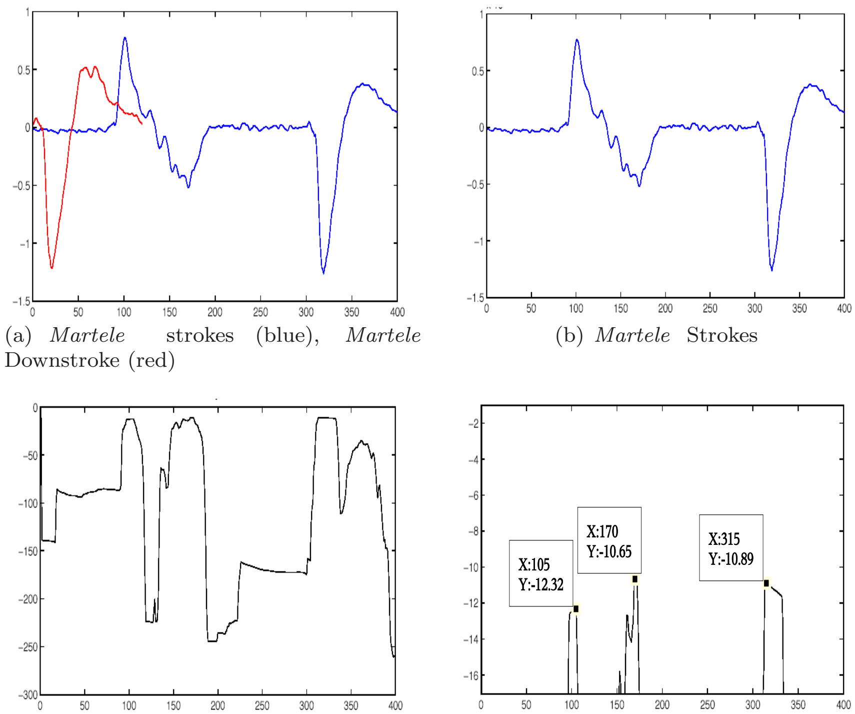
· Likelihood is equal to $y = -12$, for the opposite stroke

· Likelihood is equal to $y = -7$, for the corresponding stroke.

Differencies between values are smaller than for *Détaché*. It is due to the biphasic shape of the curve. The HMM is still recognize the shape. HMM takes in account the past evolution of the signal. It makes the difference between an acceleration curve witch started to rise and an acceleration curve witch started to drop. Permutation between positive and negative part of a *Spiccato* stroke reduces variance of likelihood.

6.1.6 Figures: *Martelé* Strokes

Figures 6.3(a) ,6.3(b) shows the tested sequence. Figure 6.3(c) shows the Likelihood, and figure 6.3(d) a zoom on Likelihood.



(a) *Martele* strokes (blue), *Martele* Downstroke (red)

(b) *Martele* Strokes

(c) Likelihood

(d) Zoom of likelihood

Figure 6.3: Likelihood for *Martele* Strokes

6.1.7 Results: *Martelé* strokes

- Likelihood is equal to $y = -12$, for the positive part of the opposite stroke
- Likelihood is equal to $y = -10.65$, for the negative part of the opposite stroke.
- Likelihood is equal to $y = -10.89$, for the corresponding stroke.

The last two values: -10.65, -10.89 are similar. HMM do not recognize the stroke.

If we look closer at the curve, shape of the curve near the local maximum is different in the two situations. For the corresponding stroke, we see a constant and slow decreasing. For the opposite stroke, we see a sharp decreasing. This difference between the two shapes let us conclude that HMM recognize the stroke.

Important Remark

Context of the sequence tested here is highly influencing the results. The constant decreasing is present because no stroke occured in the following of the sequence. The situation will be different for a fast succession of strokes. Such particular shapes in likelihood curve would no longer occure, if strokes are played faster.

6.2 Conclusion: Recognition of same bowing styles

→ HMM recognizes the good stroke in *Spiccato* and in *Détaché* bowing styles.

→ Biphasic shapes are recognized in *Spiccato* bowing style.

→ *Martele* bowing style reveals a problem. A maximum superior than the maximum expected for a stroke supposed to be recognized occured. Local maximum must be considered in a larger context. If we look at the zoom of likelihood, after the maximum for the corresponding stroke, the likelihood drops constantly and slowly. This information is necessary to conclude for the recognition of the stroke.

→ Once likelihood started to increase, we observe that it drops rapidly and then increase again. Such variations are due to the biphasic nature of strokes.

Deep analyse of figure 6.3(d) (Zoom of lilkelihood for *Martele* strokes) will explain our purpose. The likelihood increases when the second positive part of the reference matches with the first positive part of the tested stroke. Once two strokes are in quadrature, likelihood drops significantly. Then the firtst negative part of the reference corresponds to the second negative part of the tested stroke. Phase to quadrature transitions are clearly defined.

→ 'Constant' steps before increases of likelihood is a complementary information. During such periods, the reference is not matching with the tested sequence. When 'something' happens in the evolution, we should look closer at values of likelihood.

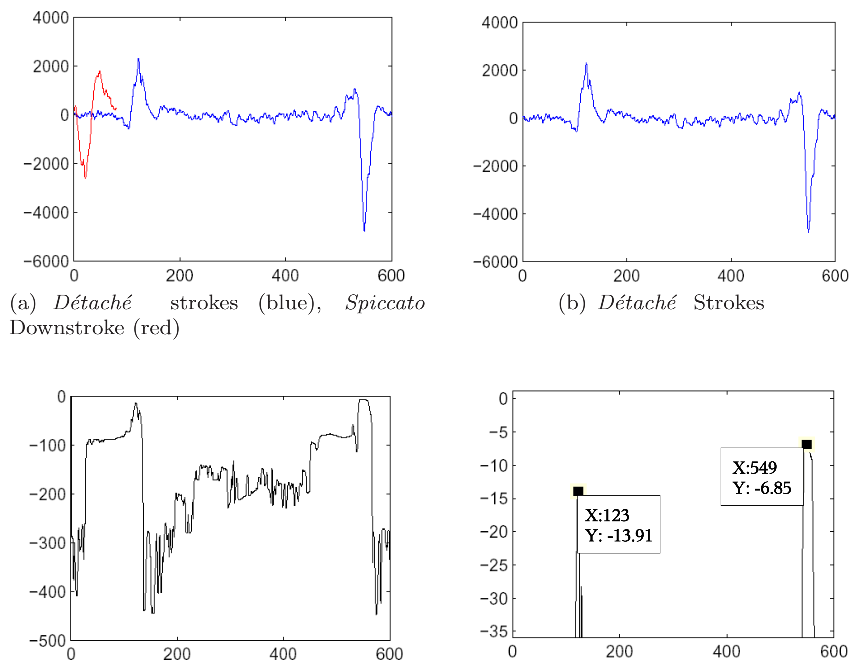
→ The more likelihood increases the more the HMM recognizes the stroke. Considering this, we assume that the *Martele* stroke is recognized.

6.3 Sequence of different bowing styles

In experiments, the sequence is one of the three bowing style. The sequence is compared to the two others bowing styles. When the sequence is a succesion of *Martelé*, reference are successively a *Detaché* and a *Spiccato* downstroke. The same tests are made for *Spiccato* and *Detaché*.

6.3.1 *Detaché* compared to *Spiccato* and *Martelé*

Figures 6.4(a) , 6.4(b) shows the tested sequence. Figure 6.4(c) shows the likelihood, and figure 6.4(d) a zoom of the Likelihood.



(a) *Détaché* strokes (blue), *Spiccato* Downstroke (red)

(b) *Détaché* Strokes

(c) Likelihood

(d) Zoom of likelihood

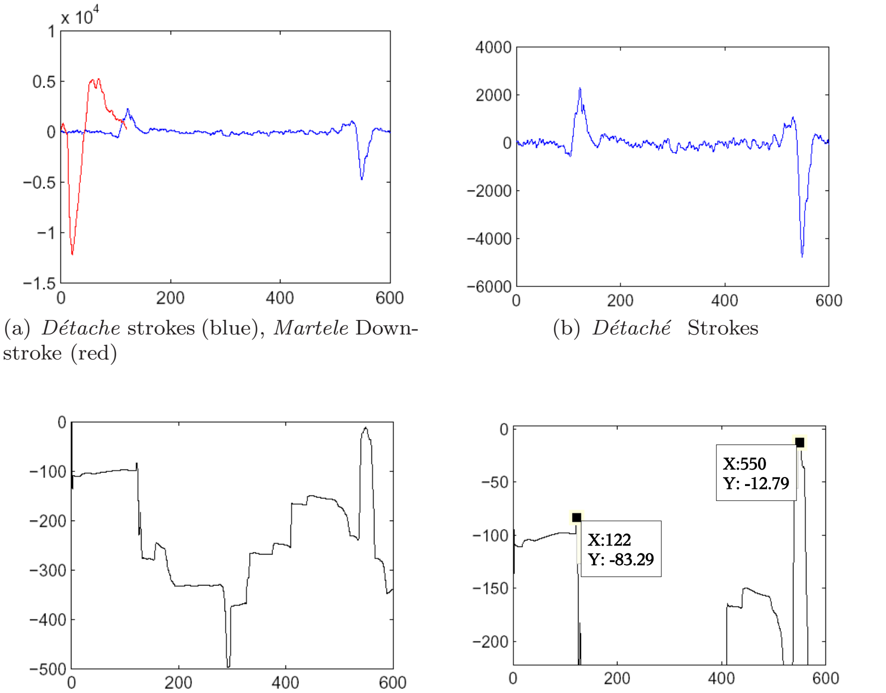
Figure 6.4: Likelihood for *Détaché* Strokes compared with *Spiccato* Downstroke

Results: *Détaché* / *Spiccato*

- Likelihood is equal to $y = -14$ (-32 [DD][1] / -12 [SS]) , for the opposite stroke
- Likelihood is equal to $y = -6.8$ (-5 [DD] / -7 [SS]) , for the corresponding stroke.

[1]The two letters in brakets follow the value of likelihood obtained when the test and the reference were the same. Example: [DD] = *Detaché* (sequence) - *Détaché* (reference).

Figures 6.5(a) , 6.5(b) shows the tested sequence. Figure 6.5(c) shows the likelihood, and figure 6.5(d) a zoom of the Likelihood.



(a) *Détache* strokes (blue), *Martele* Downstroke (red)

(b) *Détaché* Strokes

(c) Likelihood

(d) Zoom of likelihood

Figure 6.5: Likelihood for *Détaché* Strokes compared with *Martelé* Downstroke

Results: *Détaché* / *Martelé*

- Likelihood is equal to $y = -83$ (-32 [DD] / -12 [MM]) , for the opposite stroke

- Likelihood is equal to $y = -13$ (-5 [DD] / -10 [MM]) , for the corresponding stroke.

6.3.2 Conclusion: *Détaché* compared to *Spiccato* and *Martelé*

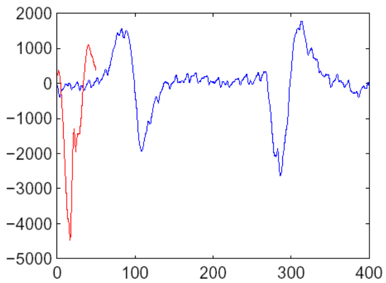
→ Values of Likelihood are lower than values obtained when the Reference and the Test were the same bowing style.

→ The *Détaché* stroke is closer to *Spiccato* than *Martele*. The maximum value of likelihood is obtained for the *Spiccato* stroke.

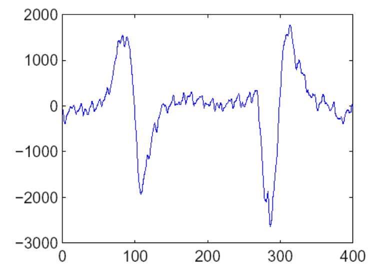
→ The direction of the stroke is recognized.

6.3.3 *Spiccato* compared to *Détaché* and *Martelé*

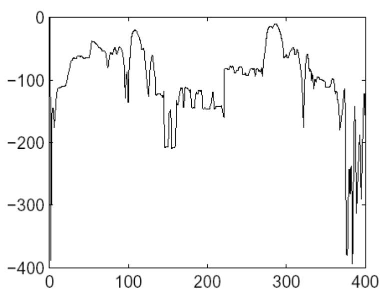
Figures 6.6(a) , 6.6(b) shows the tested sequence. Figure 6.6(c) shows the likelihood, and figure 6.6(d) a zoom of the Likelihood.



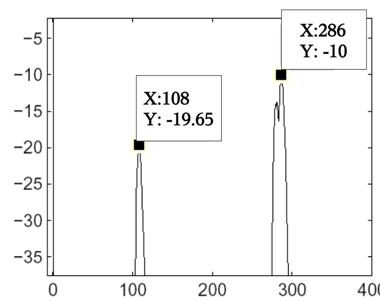
(a) *Spiccato* strokes (blue), *Détaché* Downstroke (red)



(b) *Spiccato* Strokes



(c) Likelihood



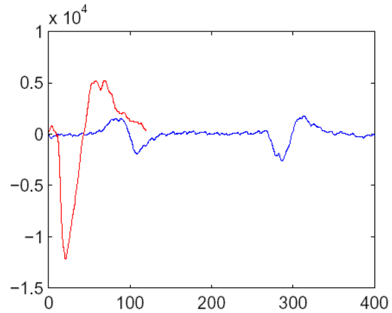
(d) Zoom of likelihood

Figure 6.6: Likelihood for *Spiccato* Strokes compared with *Détaché* Downstroke

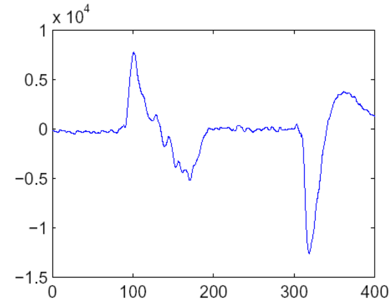
Results: *Spiccato* / *Détaché*

- Likelihood is equal to $y = -19$ (-32 [DD] / -12 [SS]) , for the opposite stroke
- Likelihood is equal to $y = -10$ (-5 [DD] / -7 [SS]) , for the corresponding stroke.

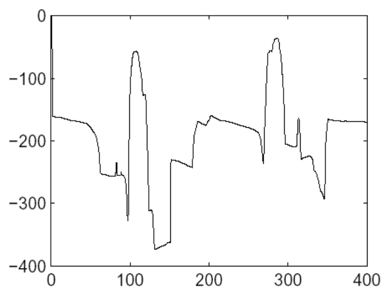
Figures 6.7(a) , 6.7(b) shows the tested sequence. Figure 6.7(c) shows the likelihood, and figure 6.7(d) a zoom of the Likelihood.



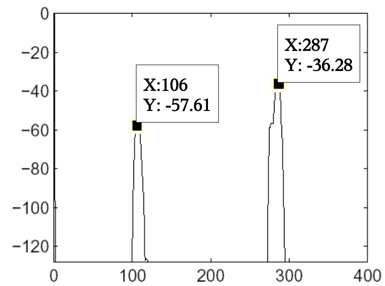
(a) *Spiccato* strokes (blue), *Martelé* Downstroke (red)



(b) *Spiccato* Strokes



(c) Likelihood



(d) Zoom of likelihood

Figure 6.7: Likelihood for *Spiccato* Strokes compared with *Martelé* Downstroke

Results: *Spiccato* / *Martelé*

- Likelihood is equal to $y = -58$ (-12 [SS] / -12 [MM]) , for the opposite stroke
- Likelihood is equal to $y = -36$ (-7 [SS] / -10 [MM]) , for the corresponding stroke.

6.3.4 Conclusion: *Spiccato* compared to *Détaché* and *Martelé*

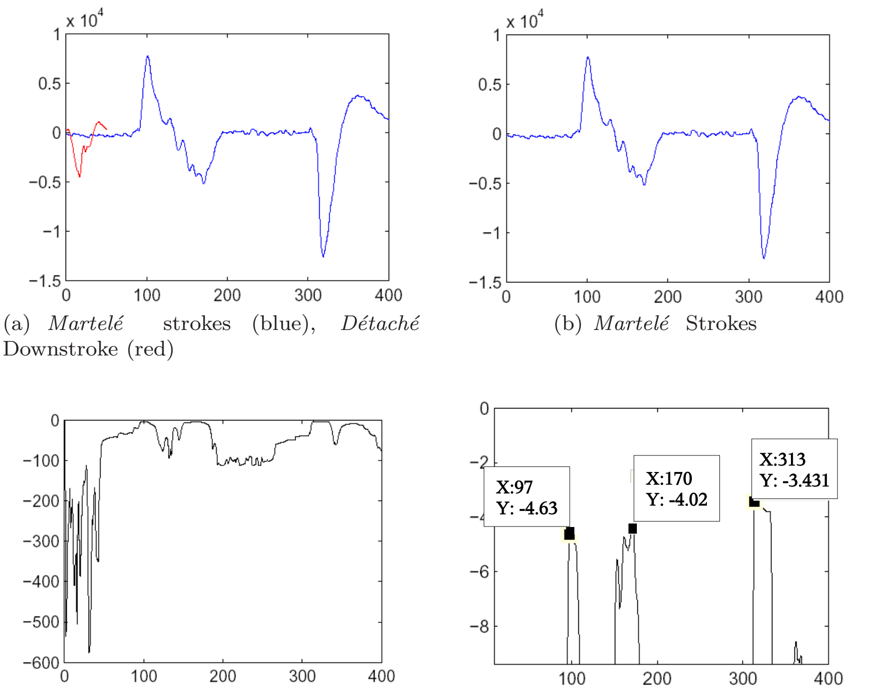
→ Values of Likelihood are lower than values obtained when the Reference and the Test were the same bowing style.

→ *Spiccato* is closer to *Detache* than *Martele*.

→ The direction of the stroke is recognized.

6.3.5 *Martelé* compared to *Détaché* and *Spiccato*

Figures 6.8(a) , 6.8(b) shows the tested sequence. Figure 6.8(c) shows the likelihood, and figure 6.8(d) a zoom of the Likelihood.



(a) *Martelé* strokes (blue), *Détaché* Downstroke (red)

(b) *Martelé* Strokes

(c) Likelihood

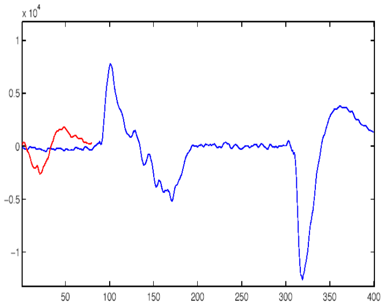
(d) Zoom of likelihood

Figure 6.8: Likelihood for *Martelé* Strokes compared with *Détaché* Downstroke

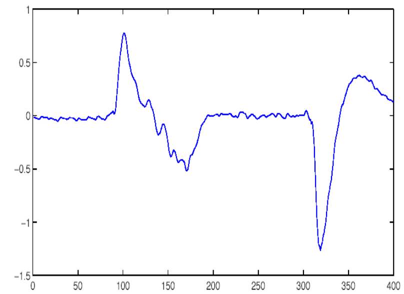
Results: *Martelé* / *Détaché*

- Likelihood is equal to $y = -4.6$ (-32 [DD] / -12 [MM]) , for the positive part of the opposite stroke.
- Likelihood is equal to $y = -4$ (-32 [DD] / -12 [MM]) , for the negative part of the opposite stroke.
- Likelihood is equal to $y = -3.4$ (-5 [DD] / -10 [MM]) , for the corresponding stroke.

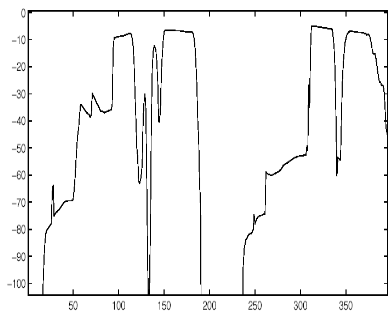
Figures 6.9(a) , 6.9(b) shows the tested sequence. Figure 6.9(c) shows the likelihood, and figure 6.9(d) a zoom of the Likelihood.



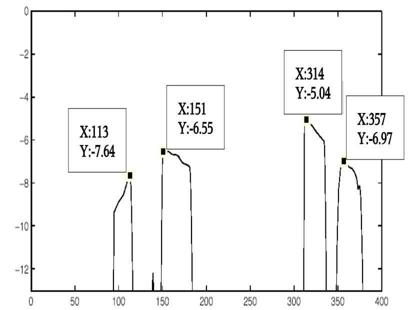
(a) *Martelé* strokes (blue), *Spiccato* Downstroke (red)



(b) *Martelé* Strokes



(c) Likelihood



(d) Zoom of likelihood

Figure 6.9: Likelihood for *Martelé* Strokes compared with *Spiccato* Downstroke

Results: *Martelé / Spiccato*

· Likelihood is equal to $y = -7.6$ (-12 [SS] / -12 [MM]) , for the positive part of the opposite stroke.

· Likelihood is equal to $y = -6.5$ (-12 [SS] / -12 [MM]) , for the negative part of the opposite stroke.

· Likelihood is equal to $y = -5$ (-7 [SS] / -10 [MM]) , for the negative part of the corresponding stroke.

· Likelihood is equal to $y = -6.9$ (-7 [SS] / -10 [MM]) , for the positive part of the corresponding stroke.

6.3.6 Conclusion: *Martelé* compared to *Détaché* and *Spiccato*

→ The direction of the stroke is recognized

→ *Martelé* is closer to *Spiccato* than *Détaché*.

→ The evolution of the curve 6.9(d) for the comparison between *Martelé* and *Spiccato* is interresting.

The two part of the *Spiccato* reference match successively with the two part of the tested strokes. The best result is obtained for the corresponding stroke. The reference is in a first time in quadrature with the tested stroke,then it is in phase. For the corresponding stroke, it is firstly in phase and then in quadrature. The typical patterns observed in likelihood curve are representative of biphasic signals.

→ Values of Likelihood are superior compared to values obtained when the Reference and the Test are the same bowing style, which might be seem counter intuitive. This fact is actually due to a limitation of AR HMM. Particularly, our procedure to set the initial variance which set to a constant through the whole stroke. This setting might be optimized through a learning algorithm.

The definition of the variance of gaussian distributions in HMM reveals itself as a difficult task. This is not a scale problem but a tolerance problem for the HMM. The question tackeled here concerns the criterion of recognition. A further analysis of the variance in HMM definition is required.

6.4 Conclusion: Recognition of different bowing styles

→ The scale of strokes is not an influencing parameter. Autoregressiv HMM solved this problem of scale.

→ The direction of stroke: Upstroke or Downstroke is always recognized.

→ Biphasic signals are partially recognized. *Spiccato* than *Détaché* are recognized. *Martelé* reveals a problem in AR Model definition.

→ Local maximum of likelihood is a good marker for recognition.

→ Particular pattern observed for biphasic signals are interresting. Transition between phase and quadrature periods is interresting.

→ Other particularities of the likelihood curve are significative:

  · Rapid increase of likelihood.
  · Constantly decrease after a local maximum.

Chapter 7

Perspectives and Future Work

Work achieved

We used *Hidden Markov Model* to recognize and segment different bowing styles: *Détaché*, *Spiccato* and *Martelé*. The data analysed are scales played on violin's strings A and E. The scales are played *mezzoforte* at a regular and relatively slow tempo.

In a first time, we used Standard HMM. The method works for *Détaché* strokes. The method does not work for *Spiccato* and *Martelé* bowing styles. It is due to the biphasic shape of these strokes. Succession of different bowing styles are problematic. Compared scales between different bowing styles is the reason of the encountered difficulties.

In a second time, we proposed an Autoregressive HMM. The direction of strokes, i.e Upstroke and Downstroke, is identified in all tests we made. The method we proposed was found efficient for the segmentation in all the cases are considered. The problem, encountered with biphasic shapes, is solved. AR HMM recognizes biphasic strokes. *Détaché* and *Spiccato* strokes respectively recognize, with more accuracy, sequences of *Détaché* and *Spiccato* strokes than others. *Martelé* sequences are problematic. Confusion occurs between *Martelé* with *Détaché* and *Spiccato*. This difficulty might be resolved by introducing a more complex learning method.

Perspectives

In all tests we made, we considered one Reference to analyse one Sequence of strokes. In a future work, a given sequence of strokes should be tested with the different bowing styles in a same operation. Gathering information, and treating them in a same process should be promissing. A raised problem is to find the best method to merge data. Computing results obtained from the different tests made, will clarify complex situations of multiple and different bowing styles played in a same sequence. Learning faculties of the HMM should be used in later work.

In our experiments, we decided to find whole strokes in sequence of bow strokes. The results we obtained are very encouraging, compared to a manual segmentation. However, the duration of the reference could be redefined. We may take a smaller reference, for example, the beginning of a stroke. The aim will be to find the smallest and the most pertinent pattern.

The reference can be, otherwise, longer. Instead of considering a single stroke, we can consider a succession of strokes. The analysis of long duration sequences, with the aim of finding a specific sequence of strokes, can be interresting for score following tasks. The analysis of long sequences may require a new model for HMM.

Once we get the segmented sequence with pattern location, we can extract gesture information to control sound synthesis, video animation or any kind of interactive installation.

We took reference's strokes from the datas we tested. The definition of reference is a delicate operation. Variability of gestures between to musical performances from a same violonist introduces new questions. HMM is supposed to recognize precise shapes. AR HMM brings the model robust to magnitude variations. Magnitude variantions are estimated in AR HMM coefficients. Such information is crucial and could be saved as a new parameter.

The influence of playing conditions such as nuance and tempo should be studied. The particular situation we studied reduced Bowstroke variability. The most influencing parameter seems to be the tempo. We observed that the shape is consequently modified when the tempo is faster. But AR HMM seems to be robust to tempo. Further experiments shall be done in this direction.

Reflexion

Gesture segmentation represents a large investigation's domain. Knowledges gathered thanks to acquisition systems will change the prime role of gesture. The gesture remained unknown for a long time. Gesture was only though to be as an ergodic process. Interactions between human and computers, at the beginning of musical computing, was mainly considered like this. Extract informations from gestures, and map them to others media is very interesting. We may access to the two other functions of the gesture. The epistemic one may be supported by the augmented reality. Perception of our environement is deeply modified. The semiotic function of gesture is potentially accessible. Gesture analysis achieved with objective tools, mapping between 'psychological' aspects of musical play and physical phenomenons will serve the semiotic of the gesture. Gesture comprehension and analysis seem to be promissing, in real time applications where causes and consequences of an human - computer interaction are naturally linked.

Appendix A

Description of the Augmented Violin

The augmented violin consist in a real violin with a bow on which we placed sensors. Sensors measure accelaration in cartesian directions X,Y and Z. Positions along bow's hairs and between bow and bridge are measured too. The system used in our research was developped at IRCAM, in 2004 by Emmanuel Flety and al. It differs from the one developped by J.Paradiso, T.Machover and D.Young. Datas are not transmit by an USB link to the workstation, they are transmit via Ethernet thanks to the *Ethersense System*. This sensor acquisition system was built by E.Fléty and al.

A.1 Measure of Positions

The first is the distance between the tip and the frog of the bow where the string is played. The second distance is the distance between the bow and the bridge, see Figure A.1.

The sensor used is a electromagnetic sensor. The system is based on the capacity coupling phenomenon to extract positions. A magnetic ribon from a videotape, used as a resistive material, covers the stick of the bow. The second element is a square-shaped antenna placed behind the bridge of the violin. Two electric signals are sent, one to the tip, and one to the frog. Those two signals are sent at different frequencies (50 and 100Hz), in order to separate them during the signal analysing process. The distance is measured through the attenuation of the signal along the stick.The more it is attenuated, the farther from the tip or the frog it is. Figure A.2.

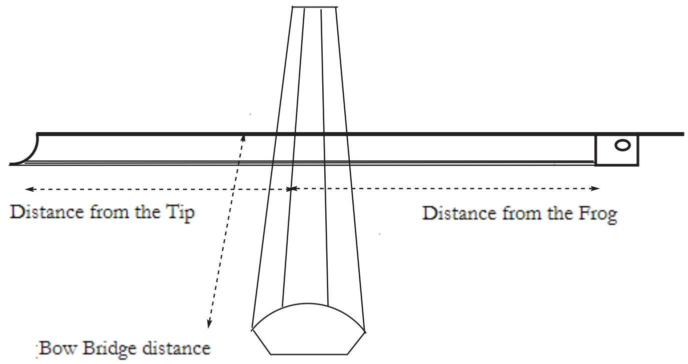


Figure A.1: Distances measured with Position Sensors

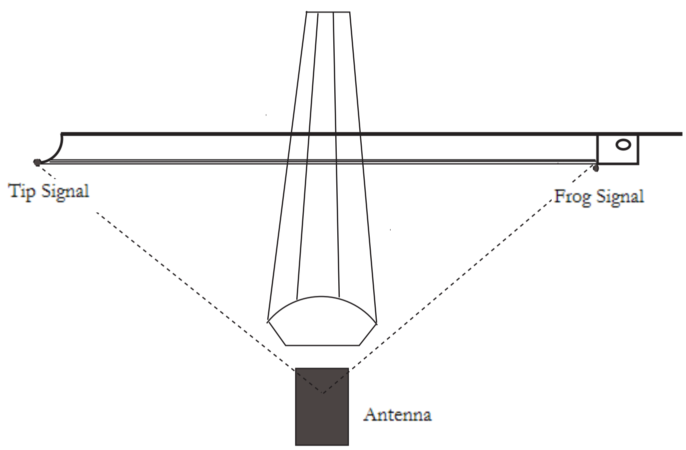


Figure A.2: Signals sent to antenna

Signals from the tip and the frog give two positions. the first is the bow bridge distance (A.1). The second is the distance where the string is played (A.2).

$$\text{Bow Bridge distance} = \frac{tip - frog}{tip + frog} \tag{A.1}$$

$$\text{Position} = \frac{1}{tip + frog} \tag{A.2}$$

Figure A.3: The augmented Bow

Figure A.4: The Bridge Antenna

A.2 Measure of Accelerations

The system used is a mass-spring system. One system is a dual axis accelerometer, the other is a one axis. The differences of bow acceleration along a direction are transmitted to mass in the accelerometer. The greater the difference is the more mass move. Friction of the mass in accelerometer is compensated afterwards by an offset correction.

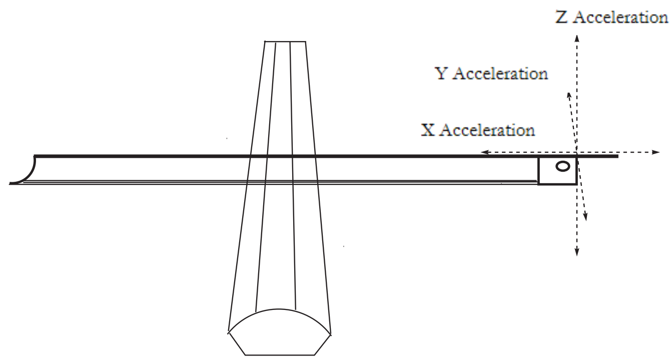


Figure A.5: Accelerations measured with Accelerometers

Figure A.6: The 3D accelometer

Appendix B

The use of Hidden Markov Model

B.1 Complexity of temporal Processes

A real measured signal of a natural process differs from one record to another. This variability of measurements can be reduced by using statistic methods.

The traditional use of statistic tools such as variance, covariance, correlation and so on, is limited. The two observations must be similar in term of magnitude and time scale. Those methods can be added to filter wich may substract noise in the signal. Stochastic distribution of signal shape is one of the major difficulty encountered in gesture movement. We may recognize signal as a class or a family of signals.

Once we introduce classes of signals, the problem of recognition is still not solved. How many classes are necessary to represent all signals wich can by observed. Thus, the study of similarity may be interesting as a first recognition process. The next step may be achieved by using the Hidden Markov Model.

Works mentionned in [38] and in [39], conform us in the use of HMM. In [38], the author describes a method to recognize shapes. In [39], the author used a advanced method called Paramatric HMM to recognize the human gesture. Such an approach, gave us the idea to use HMM for our segmentation problem.

B.2 Hidden Markov Model

After a preliminary work in wich we studied the similarities between two signals, it appeared necessary to use the Hidden Markov Models. Here we will explain with more details what we developped in section (2.1) page 5.

Hidden Markov Model is considering the processus as a set of N states. The processus evolution is described by transitions between states. The elements constituting the HMM are:

- A set of states S. $S = \{S_1, S_2, ..., S_N\}$

- A state transition probability distribution, called *Transition Matrix* (fig(B.1)). $A = \{a_{ij}\}$, representing the probability to go from state S_i to S_j

  $a_{ij} = P[q_{t+1} = S_j | q_t = S_i] \quad 1 \leq i, j \leq N, \quad a_{ij} \geq 0, \quad \sum_{j=1}^{N} a_{ij} = 1$

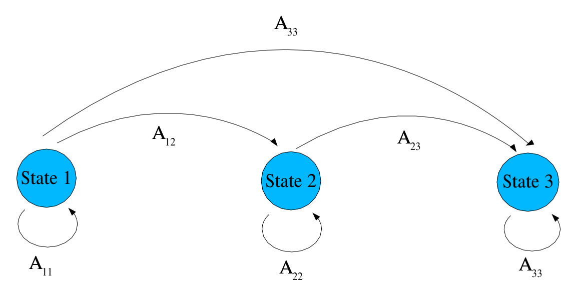


Figure B.1: Illustration of the Transition Matrix

- A set of observations V. $V = \{v_1, v_2, ..., v_M\}$

- An observation symbol probability distribution, called *Emission matrix* (fig(B.2)). $B = \{b_j(k)\}$, representing the probability of emission of symbol $\{v_k\}$ when system state is $\{S_j\}$.

$$b_j(k) = P[v_k(t)|q_t = S_j] \quad 1 \leq i, j \leq M, \quad b_j(k) \geq 0, \quad \sum_{j=1}^{M} b_j(k) = 1$$

- An initial state probability distribution $\pi = \{\pi_i\}$, representing probabilities of initials states.

  $\pi_i = P[q_1 = S_i] \quad 1 \leq N, \quad \pi_i \geq 0, \quad \sum_{i=1}^{N} \pi_i = 1$

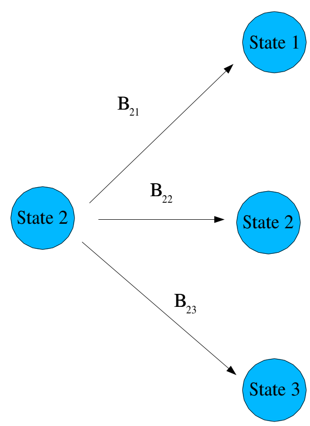


Figure B.2: Illustration of the Emission Matrix

B.3 Particularities of the Model

B.3.1 Specific Transition Matrix

We are studying temporal processes. It means, in term of direction for transitions, that it goes from left to right. The transition matrix is however, particular. Transition matrix coefficients are null except:

$a_{k,k} = 1 - \frac{1}{d}, \quad a_{k,k+1} = 1 - a_{k,k}, \quad 1 \leq k \leq N - 1$

$d =$ distance between two successive states

Auto transition $a_{k,k}$ are the dominant transitions. Transition from state S_k to S_{k+1} is the only other transition allowed. Distance between states is very influencing.

- If d is equal to 2, system has equal chance to stay in state S_k or to go to state S_{k+1}.

- If d is greater than 2, system has more chance to stay in state S_k.

B.3.2 Specific Emmission Matrix

The probabilities of emission of symbols v_k for the state S_j are described with a gaussian distribution. For a given state, we consider around its location a gaussian curve (fig(B.3)). It brings to the HMM a certain flexibility.

Definition of a gaussian distribution:

$$\mathrm{N}(\mu,\sigma) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

In our Model:

$$b_j(k) = P[v_k(t)|S_j] = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_k-\mu)^2}{2\sigma^2}}, \qquad \mu = \frac{S_j + S_{j+1}}{2}, \qquad \sigma = cst$$

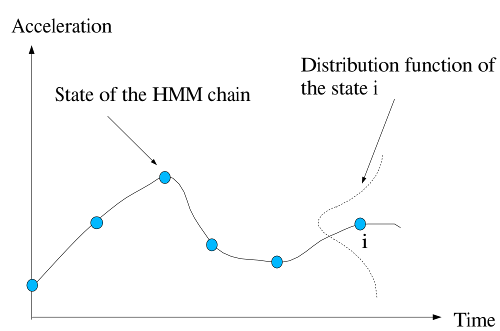


Figure B.3: The HMM reference

B.4 The recognition process

Sample the Observation Curve

To model the acceleration curve, the first step consist in sub sample the curve to define the states of the HMM (fig(B.4)).

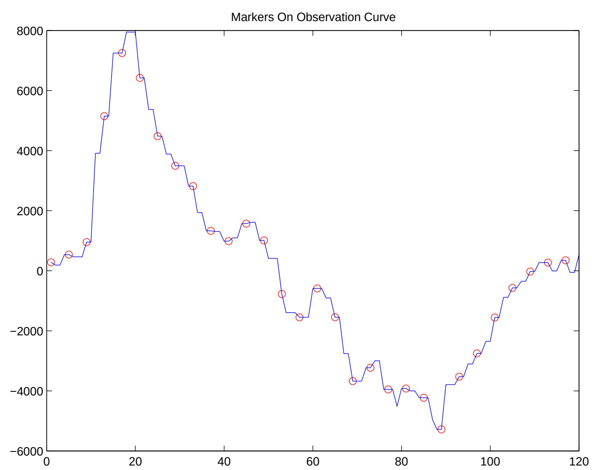


Figure B.4: Sub sample of the Curve

From States Sequence to Transition Matrix

After this operation we are disposing of the A matrix, the Transition matrix.

Emission Matrix

$B = \{b_j(k)\}$, the Emission Matrix, is defined here. The number of symbols v_k is equal to N the number of states. With all considerations made before, the coefficient of B are:

$$b_j(k) = \frac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(v_k - \mu(S_j, S_{j+1}))^2}{2\sigma^2}}, \qquad 1 \leq k \leq N, \qquad 1 \leq j \leq T. \quad \text{(B.1)}$$

$$\mu(S_j, S_{j+1}) = \frac{S_j + S_{j+1}}{2}, \qquad T = \text{length of the observation curve} \quad \text{(B.2)}$$

Estimation of likelihood

1. The estimation Matrix is a [T,N] matrix initialized at $\alpha_{1,1} = 1$

2. The first operation consist in multiplying Estimation Matrix $Alpha = \{\alpha_{j,k}\}$ with Transtion Matrix $A = \{a_{k,k}\}$. Because $k = nN$, matrix dimensions agree. New values expression is:

   $\alpha_{j,k} = a_{k,k} \cdot \alpha_{j,k}$

3. Second Step: Each state S_j of the observation is compared to all possible reference's symbols. We introduce a new Matrix Bj.(fig(B.5))

   The new $Bj = \{b^j_{k,j}\}$ matrix coefficients are products of the current state with the Emission Matrix B.

   $b^j_{j,k} = S_j \cdot b_{j,k}$

   It means that the most probable symbol is straight. The others have less weight in the calcul of probability.Then:

   $\alpha_{j,k} = b^j_{k,j} \cdot \alpha_{j,k}$

4. Third step: To estimate the global likelihood between an Observation and the Reference, we sum rows of Estimation Matrix $Alpha = \{\alpha_{j,k}\}$.

   $x_j = \sum_{k=1}^{N} \alpha_{j,k}$

   For an Observation on a T duration, The likelihood with the Reference is estimated. For computing reasons, we are studying $log(x)$

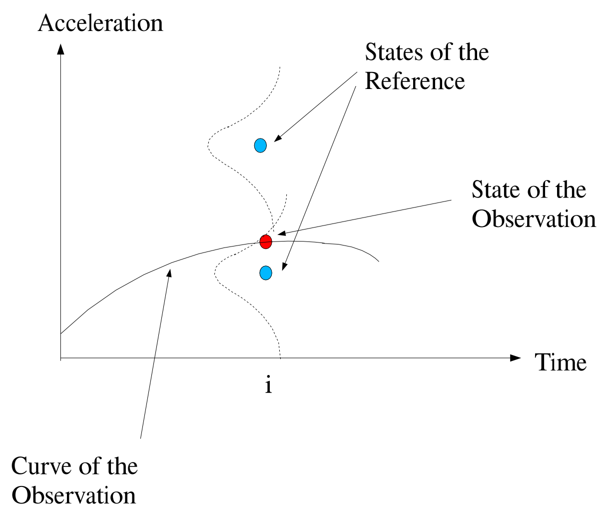


Figure B.5: The comparison process

Bibliography

[1] Claude Cadoz. Instrumental gesture and musical composition. In *Proceedings of the International Computer Music Conference, ICMC*.

[2] Claude Cadoz. Les nouveaux gestes de la musique. In *Musique, geste, technologie*.

[3] M.M. Wanderley and M. Battier, editors. *Trends in Gestural Control of Music*. Ircam, 2000.

[4] Marcelo Wanderley. *Performer-Instrument interaction : applications to gestural control of sound synthesis*. Thèse, Université Paris 6, 2001.

[5] Emily Morin. Captation de modes de jeu instrumentaux : le cas du violoncelle. Technical report, Ircam, 1999.

[6] Emily Morin. Analyse des coups d'archet du violoncelliste avec le dispositif digibow. Technical report, Ircam, 2000.

[7] Emmanuel Fléty, Nicolas Leroy, Jean-Christophe Ravarini, and Frédéric Bevilacqua. Versatile sensor acquisition system utilizing network technology. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.

[8] N.Rasamimanana, Emmanuel Fléty, and Frédéric Bevilacqua. Gesture analysis of violin bow strokes, 2004.

[9] Bernd Schoner, Chuck Cooper, Chris Douglas, and Neil Gershenfeld. Cluster-weighted sampling for synthesis and cross-synthesis of violin family instrument. In *Proceedings of the International Computer Music Conference, ICMC*, 2000.

[10] Lawrence R.Rabiner. A tutorial on hidden markov models and selected applications in speech recognition. 1989.

[11] Knut Guettler. Looking at starting transients and tone coloring of the bowed string. 2004.

[12] Anders Askenfelt. Measurement of the bowing parameters in violin playing. ii: Bow-bridge distance, dynamic range, and limits of bow force. *J. Acoust. Soc. Am.*, 86(2), 1989.

[13] Beaudoin-Lafon. Interaction instrumentale: de la manipulation directe à la réalité augmentée. In *Actes Neuvièmes Journées sur l' Ingénierie des Interfaces Homme-Machine*.

[14] Dan Trueman and Perry R. Cook. Bossa: The deconstructed violin reconstructed. In *Proceedings of the International Computer Music Conference, ICMC*, 1999.

[15] Joseph A. Paradiso and Neil Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, 21(2), 1997.

[16] Diana Young. The hyperbow controller: Real-time dynamics measurement of violin performance. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, Dublin, Ireland, 2002.

[17] Diana Young. Wireless sensor system for measurement of violin bowing parameters. In *Proceedings of the Stockholm Music Acoustics Conference, SMAC*, 2003.

[18] Stefania Serafin and Diana Young. Bowed string physical model validation through use of a bow controller and examination of bow strokes. In *Proceedings of the Stockholm Music Acoustics Conference, SMAC*, 2003.

[19] Paul Brossier, Juan Bello, and Mark D.Plumbley. Real time temporal segmentation of note objects in music signals. 2004.

[20] Mark Billinghurst. Gestured based interaction. 2002.

[21] Claude Cadoz and Marcelo M. Wanderley. Gesture music. *Trends in Gestural Control of Music*, 2000.

[22] Chad Peiper, David Warden, and Guy Garnett. An interface for real-time classification of articulations produced by violin bowing. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2003.

[23] A. Camurri, B. Mazzarino, M. Richetti, R. Timmers, and G. Volpe. Multimodal analysis of expressive gesture in music and dance performances. 2003.

[24] Harling and A. D. N. Edwards. Hand tension as gesture segmentation cues. In *Proceedings of Gesture Workshop'96*.

[25] Hofmann, F., and G. Hommel. Analyzing human gesture motions using acceleration sensors. In *Proceedings of Gesture Workshop'96*.

[26] Kristoffer Jensen. Musical instruments control and expression.

[27] Stefania Serafin, Frederic Mazzella, Julius O. Smith III Arnaud Tellier, and Harvey Thornburg Guillaume Thonier. Data driven identification and computer animation of bowed string model.

[28] Bernd Schoner, Chuck Cooper, Chris Douglas, and Neil Gershenfeld. Data driven modelling and synthesis of musical instruments.

[29] Sylviane Sapir. Interactive digital audio environments: Gesture as a musical parameter. In *Proceedings of the COST G-6 Conference on Digital Audio Effects*, 2000.

[30] Ben Nevile, Peter Driessen, and W. A. Schloss. A new control paradigm: Software-based gesture analysis for music.

[31] Frdric Bevilacqua, Christopher Dobrian, and Jeff Ridenour. Mapping sound to human gesture: demos from video-based motion capture systems. 2000.

[32] Jean-Michel Couturier. A scanned synthesis virtual instrument.

[33] Suguru Goto Takahiko Suzuki. The case study of application of advanced gesture interface and mapping interface - virtual musical instrument "le superpolm" and gesture controller "bodysuit". In *In Proccedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.

[34] Curtis Bahn and Dan Trueman. Interface electronic chamber ensemble. 2000.

[35] Analytic Chemistry, editor. *Smoothing and Differentiation of Data by Simplified Least Squares Procedures*. 1964.

[36] Gidon Eshel. The yule walker equations for the ar coefficients. 2004.

[37] Antonello Panuccio, Manuele Bicego, and Vittorio Murino. A hidden markov model-based approach to sequential data clustering. 2005.

[38] Xianping Ge and Padhraic Smyth. Deformable markov model temtaples for time series pattern matching. 2000.

[39] David Wilson. Adaptative models for the recognition of human gesture. 2000.