Gesture Analysis of Bow Strokes Using an Augmented Violin

Nicolas Hainiandry Rasamimanana
Mémoire de stage de DEA ATIAM année 2003-2004
Université Pierre et Marie Curie, Paris VI

Laboratoire d'accueil : Ircam - Applications Temps Réel
Responsable : Frédéric Bevilacqua

Mars 2003 - Juin 2004

Contents

List of Figures

List of Tables

Abstract

At Ircam, we are currently developing an augmented violin. Such an augmented instrument appears to be attractive in computer music because it offers a large diversity of sounds and nuances. In order to study how this ability can be used to have a subtle, continuous control on musical processes (e.g sound synthesis), we have placed a gesture sensing system on a violin carbon fiber bow. With the help of this equipment, we have particularly analyzed the bow speed, which is one of the most influencing parameter on sound in the playing of a bowed string instrument, according to instrument acoustics. In this study, we have notably focused on four different types of bow strokes: détaché, martelé, piqué and spiccato. Their gestural data analysis has resulted in the constitution of a set of features corresponding to the interest points in the speed curve, i.e extrema and inflection points. We have tested these features on a bow stroke database that we built using our segmentation algorithm on recording measurements. The features show some strong invariance properties for a single violinist and between two different violin players including Jeanne-Marie Conquer. The features behavior is also pertinent according to gesture variations, especially when changing nuances and tempo. Moreover, this feature space is consistent with acoustics studies having shown the influence of bow speed on sound spectral characteristics: the features being extracted from the speed temporal curve, we can deduce some perceptual properties of the space generated by the features.

Résumé

Nous sommes actuellement en train de développer un violon augmenté à l'Ircam. Un tel instrument suscite un attrait particulier dans le domaine de l'informatique musicale parce qu'il offre un large éventail de sons et de nuances. De manière à étudier comment cette particularité peut être utilisée pour contrôler finement et de manière continue des processus musicaux, nous avons placé un système de captation du geste sur un archet de violon en fibre de carbone. A l'aide de ce système, nous avons en particulier étudié la vitesse de l'archet, l'un des paramètres les plus influents sur le son dans le jeu d'un instrument à cordes frottées, d'après l'acoustique instrumentale. Dans cette étude, nous nous sommes notament intéressés à quatre types différents de coups d'archet : détaché, martelé, piqué et spiccato. L'analyse de leur données gestuelles a permis de constituer un ensemble de descripteurs correspondant aux points d'intérêt de la courbe de vitesse, i.e les extrema et les points d'inflexion. Nous avons testé ces descripteurs sur une base de données de coups d'archet que nous avons construite en utilisant notre algorithme de segmentation sur des enregistrements de mesures. Les descripteurs montrent de fortes propriétés d'invariance pour un et deux violonistes (dont Jeanne-Marie Conquer). Le comportement de ces descripteurs est également pertinent en terme de variations sur les gestes, en particulier dans les changements de nuances et de tempo. De plus, l'espace engendré par ces descripteurs est compatible avec les résultats d'études acoustiques montrant l'influence de la vitesse d'archet sur les propriétés spectrales du son : les descripteurs gestuels étant extraits de la courbe temporelle de la vitesse, on peut en déduire des propriétés perceptuelles pour l'espace engendré par ces descripteurs.

Acknowledgments

I would like to acknowledge:

Norbert Schnell, who welcomed me in the Applications Temps Réel research team,

Frédéric Bevilacqua, my supervisor, for the fruitful discussions we had,

Emmanuel Flety, for the tremendous work he did to build the augmented violin prototype,

Alain Terrier, who made the carbon clip we use to fix the electronics boards on the bow,

Diemo, for always answering my questions on Matlab and LaTeX.

Gérard Assayag and Cyrille Defaye, for their availability throughout the year.

I would also like to acknowledge the artists who contribute to the project, in particular the composers Philippe Manoury, Florence Baschet, Franck Bedrossian and Jerome Combier and address a special thank to professional violinists Hae Sun Kang and Jeanne-Marie Conquer for their gesture recordings.

And last but not least, I would like to acknowledge all my classmates from DEA ATIAM for their kindness, with a special note to Arshia, my office-mate, who was forced to bear me more than the others. Arshia, you are right, we had a nice time.

Introduction

This work was done in the context of *Ircam*'s interest in movement analysis. The objective is to find some different means of interaction with a computer that would involve gestures. Here, we focus on the study of the violin bowing techniques. The subtleties induced in the playing of a bowed instrument should lead to the creation of a pecularly rich control interface. This research project is also motivated by a group of composers, including notably Philippe Manoury, Florence Baschet, Franck Bedrossian and Jerome Combier, who intend to use the results in future pieces and who, as a consequence, actively participate in its development.

The gesture analysis is performed via a sensing system mounted on a carbon fiber violin bow. A number of various signals, ranging from the bow position to the downward force on the strings, are then issued and taken as input. There are two distinct ways of using these continuous signals. The first one consists in a direct use of the data with a minimal interpretation: it can then serve as parameters to control a physical model synthesis, e.g. that of a bowed string. The other way of using the input tries to interpret the signals in order to determine some higher level information, e.g. the performed bow stroke is staccato.

The former way of use is rather straightforward in that the main difficulty resides in feeding a physical model synthesis module with the right parameters, while the latter demands deeper analysis but provides a better understanding of the sensor signals behavior. Musical applications can therefore take advantage of this knowledge and for example follow the interpret gesture in the same way as score following do with audio, which is of main interest to trigger events in mixed pieces, or anticipate the violinist movements.

These two applications pose the problem of identifying objects in the violinist gesture and in the sensor signals. In the same manner that score following relies on tones, which characteristics are notably pitch and duration, we here have to determine what is to follow or to anticipate. Therefore, from these continuous sensor signals, we have to extract pertinent pieces of information on the bowing. In other words, it would be interesting to eventually try to identify patterns and invariants.

The work done here focuses on the study of a set of different bow strokes in order to characterize some playing aspects of the violin techniques. Using the sensing system described in chapter 2, we recorded strokes gestures to analyze them. After discussing the accuracy of the sensing in system in chapter 4, we will present the different features extracted from the gestural data in chapter 5 and show their invariant properties in chapter 6.

Chapter 1

State of the art

1.1 Introduction

A music instrument can be defined as the meeting between art and technology. However, technology and more pecularly computer science has known an exponential growth during these last fifty years, and so the desire of creating new instruments that would exploit these developments. These instruments, often called Digital Music Instrument (DMI), can be divided into two categories: the new music interfaces and the augmented instruments.

The first category involves new controllers made from scratch using various sensors, e.g. accelerometers or force sensing resistors, and use the gestural data with a low level interpretation by mapping them to sounds. This approach is particularly attractive as it generally offers a straightforward interface. However, the problem of a long and complex learning found with a traditional instrument is often replaced by the searching of appropriate mappings between gesture and sound, as these correspondences are totally opened, which can be seen as hard a task to do. Moreover, these new interface simplicity turns out to be one of their main drawbacks too. It indeed often means a poorer expressive interface. Another important point is the presence of a haptic feedback, which plays an important role in the relationship between an interpret and its instrument. Indeed, the consideration of the physical response when a musician performs a gesture is a determinative element in the mastering of his/her instrument. In this first approach, this aspect is still under development at the moment in order to offer something comparable to the feeling of an instrument. However, there are already a number of very interesting and fascinating works, like the worth seeing on stage performances of Atau Tanaka with the BioMuse [18]. Bioelectrical signals, and particularly electromyograms of his arm muscles, are digitized and mapped to sounds and images. Therefore, the movements of his body are directly interpreted to create music. Atau Tanaka underlines that although BioMuse is not a mechanical instrument because there is no material object manipulated by the musician, it is quite physical as it depends directly on corporeal gesture [18].

The second category is the one we chose and is based on the adjunction of a sensing system on an acoustic instrument. The idea here is to use a traditional instrument and to extend its qualities by electronic means therefore creating a new instrument. The main advantage of this approach is that this time the instrument already comes with a large set of playing techniques which reflects many possibilities of expressiveness. However, it is also its main drawback because the possible transformations that can be brought to the instrument are largely conditioned by the instrument history, repertoire and previous techniques of play. The interesting side of this approach is to see to what extent electronics can contribute to, and be considered as an evolution of the instrument.

A music instrument reflects the favorite correspondence between gesture and sound, between action and perception, so that an interpret can give way to expressiveness. The two options described here have completely different starting points, but both tries to enrich expression possibilities. In this report, we focus on what valuable information can be drawn from an augmented violin, research being motivated by the violin complexity of play and diversity in timbres and nuances.

1.2 Previous works and applications

There is already a certain number of works on sensing system hardware to record the violinist bowing gesture with different applications.

In 1986, Tod Machover founded the HyperInstruments research group at the MIT Media Lab, with the aim of designing expanded instruments with electronic technology. In 1999, Diana Young built a new violin interface, the HyperBow, based on one of Tod Machover's first works on HyperInstruments, the HyperCello (1991) which was created for Yo-Yo Ma. Diana Young's violin bow sensors [22], [24], [25], [23] measure the bow position, i.e. the distance between contact point of the bow with the strings the bow frog and the distance from tip to frog with a capacity coupling system, the bow acceleration in three axis with accelerometers, and the downward and lateral forces on the bow stick with strain gauges. The gauges help measuring the flexion of the bow, which can be regarded as an image of the bow hair downward force on the strings. However, they present many implementation issues and constraints among which the problem of adhering on a cylindrical surface, the carbon heat dissipation properties versus the degradation of the gauges performances at high temperature and their fixing being permanent. The collected data are then sent wirelessly to a bay station and then via a serial/USB port to a workstation. The HyperBow has been integrated in Toy Symphony for full orchestra, children's chorus and solo violin, written by Tod Machover. In this piece, the HyperBow is used to control the activation and the alteration of sounds and effects on an electric violin, using a selection of chosen gestures. The HyperBow has also been used to evaluate the playability of various physical models of friction based instruments, including a violin [26], [16], a Tibetan singing bowl, a musical saw, a glass harmonica and a bowed cymbal [17]. Here, means to expressively control these models and involving the HyperBow are investigated.

In 1998 [13] and 2000 [14], Bernd Schoner and al uses probabilistic techniques to infer, in real-time, violin sounds from the gesture input given by the HyperCello bow, the previous version of the HyperBow. He uses a cluster-weighted modeling to predict the sound pitch and amplitude from the gestural data, and tries to extend the inference model to spectrum with rather good results in sustained part but with mitigated results in the transitions.

In 1999, Perry Cook and Dan Trueman [19] built a new instrument based on the different elements constituting a violin, the BoSSA (Bowed-Sensor-Speaker-Array). Its bow, the R-Bow, is very close to that of Tod Machover: it is a standard violin bow fitted with pressure sensors (force sensing resistors, FSRs) and a dual axis accelerometer which measures both angle position and changes of velocity. The BoSSA Fangerbored is a fingerboard augmented with a linear position sensor, four FSRs to use by the right hand to trigger events, and another dual axis accelerometer. The strings are replaced by an array of four pieces of foam-covered wood, "sponges", resting freely between two fixed FSRs and that can be bowed as real strings. What serves as a resonating body is a spherical multi-channel speaker arrays, which can reconstruct the radiative timbral qualities of violins in a traditional acoustic space. The proximity between input and output grants BoSSA with a playing similarity to an acoustic instrument with the flexibility of software synthesis and signal processing techniques. The BoSSA was used in The Lobster Quadrille, a piece composed by Dan Trueman, where the bow gesture data is used to control a comb filter vibrato.

In 2001, Camille Goudeseune with his eViolin [5] tracks the violinist movements by means of an electromagnetic field. The sensing system is composed of two sensors placed on a five-string electric violin and of an antenna emitting a time-varying magnetic field and placed about one meter across. The spatial position of the violin is mapped to timbre according to some perceptual dimensions: spectral brightness as a function of latitude and spectral richness as a function of longitude. The third spatial dimension is not mapped to a third perceptual dimension because of the inconvenience of playing at abnormal altitudes. It is rather used to toggle switch (octave switch) or as a continuous scaling factor for the amount of reverberation for example. It is to be noted that in the eViolin applications [7] the electric violin is associated to a output array of speakers similar in efficiency to that of BoSSA [19] in an apparently cheaper version.

Since 2000, Charles Nichols [10] has developed two versions of a virtual violin bow haptic human-computer interface, the vBow. The bow is custom-made of fiberglass and is linked to servomotors in order to sense the bow position and bring a haptic feedback. The data is then used to drive a physical model synthesis.

In his CyberViolin project (2003), Chad Peiper and al [11] uses an electromagnetic system to record the position of two sensors mounted on a violin bow. He offers a higher level interpretation of the gesture data as he extracts some features to classify the violin different bow strokes using a decision tree. The features he uses are the bow position at the beginning and at the end of the stroke, the bow speed, the frequency of bow change, the acceleration or deceleration within a stroke, the continuity of motion between strokes, the lack of movement within a stroke. These features are then provided to a decision tree for training and recognition. The performances of his classification process are actually limited by the inaccuracy of the sensing system (sensor errors, resolution and sampling frequency) and should be improved by additional features in the decision tree. The CyberViolin interpreted data are used in a 3D graphical environment in two ways: as a representation of the performed bow strokes, which grants the violinist with a real time feedback, and as a means to interact with a program using the bow instead of a traditional pointing device.

1.3 Ircam Prior Works

Ircam has a strong background in mixed pieces, where an acoustic instrument and a computer perform together. Philippe Manoury's piece *Jupiter* for flute and real time electronics, composed in 1987, is a pioneering work in the field of interaction between a live instrument and a computer. The 4X, which was designed by Giuseppe Di Giugno, used the audio and score following to interact with the interpret while the flute, conceived by Larry Beauregard, was augmented so that its fingering could be detected and the instrument used as a control input device.

Several studies have already been carried out at Ircam on gestural control, with some major works by Marcelo Wanderley in [20] and [21]. In 2000, Emily Morin studied the similarities and differences between different cellists' way of bowing using a DataGlove, [9] and [8]. The FSRs located on each contact points between the right hand and the bow showed some repeatable patterns that could be used in a recognition process, although the pressure of the fingers on the bow stick is not very reliable a parameter.

In 1996, Suguru Goto built *"le SuperPolm"* with the collaboration of Patrice Pierrot. *"Le SuperPolm"* is a control interface that is based on a violin but has no strings nor hair bow. It is played in a similar way to a violin so that the body movements can be recorded and used as input data to trigger events as in [4].

In this report, we will describe the latest version of the augmented violin bow developped in 2004 by Emmanuel Flety at Ircam and the analysis of the collected data.

Many works have been done to exploit the violin expressive possibilities. The major part has directly used the sensor signals to control sound treatment parameters, such as filters or physical model synthesis. Chad Peiper [11] and Bernd Schoner [13] have directly used the gestural data in theoretical frameworks (prediction and decision trees) in order to make some correspondence with sound. In this report, we propose a study on the nature of the violinist gesture and the measured signals, in order to extract their invariance and variability characteristics. This analysis leads us to a violin gesture representation space with pertinent perceptual properties.

Chapter 2

Ircam's Augmented Violin

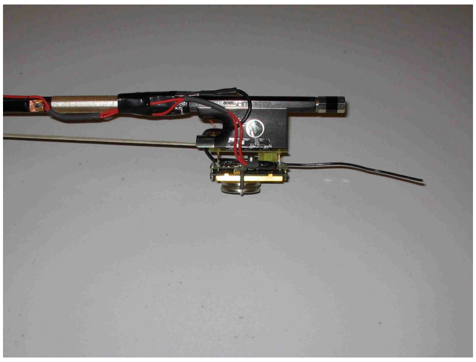
The sensing system has been built by Emmanuel Flety at Ircam in 2004. It is largely inspired by the HyperBow system, developed in M.I.T Media Lab by Joe Paradiso, Tod Machover and Diana Young [24], which is one of the lightest and most unintrusive wireless system. However, the overall architecture is rather different: the data is not sent to the workstation by a serial/USB connection but via ethernet thanks to *EtherSense*, a sensor acquisition system developed by Emmanuel Flety and al [3].

2.1 Sensing System Description

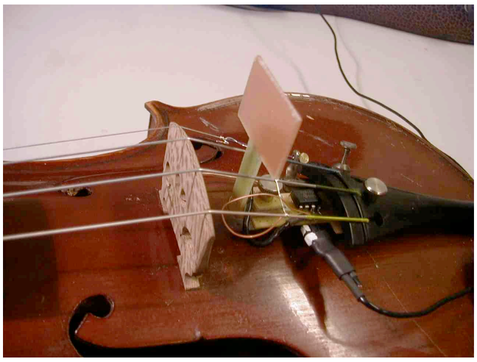
In order to measure the violinist movements in a playing situation, several sensors have been added to a conventional carbon fiber bow. The collected data include the bow position respectfully to the bridge, the bow position between tip and frog, the acceleration along three axis (in the bow stick direction, in the string direction and orthogonally to the bow in the vertical direction), the angle of the bow corresponding to its rotation respectfully to these three axis, and the forefinger pressure on the bow stick. (See picture on figure 2.1).

2.1.1 Position Sensor

The bow positions (to the bridge and from tip to frog) are deduced from the same electromagnetic position sensor. The method used is based on a capacity coupling between the bow and a square-shaped antenna placed behind the violin bridge. To do so, the bow is covered with a resistive material, here a piece of the magnetic ribbon of a video tape, that runs alength the stick. Two electric signal are sent at each extremities of the bow at different frequencies (50kHz and 100kHz) and are gradually attenuated along the bow stick by the resistance. The tip and frog signals are received in a single electrical signal and demodulated with a low-pass and a high pass



(a) Augmented violin bow



(b) Position sensor antenna

Figure 2.1: The violin sensing system. The sensing system is mounted on a carbon fiber bow. The clip under the bow contains the RF transmitter, the electronics board holding the micro-controller, the accelerometers. The FSR is placed on top of the metallic grip. The position sensor system includes the resistive tape (starting from the copper clip) and the antenna mounted behind the bridge.

filter.

The positions are then computed according to the strength of the two signals. Ideally, both electric signals should decrease linearly with the same slope along the resistance. With this assumption, the bow-bridge distance and the tip-to-frog distance can be deduced by the following equations:

$$bow\_bridge\_dist = \frac{tip - frog}{tip + frog}$$

and

$$tip\_to\_frog\_dist = \frac{1}{tip + frog}$$

where *tip* [resp. *frog*] is the strength of the signal emitted from the tip [resp. frog].

2.1.2 Acceleration Sensor

In addition to the position sensing system, two accelerometers (a dual axis and a single axis) are used in order to measure the variations of speed over time in three directions: the bow direction and the string direction by an AnalogDevice ADXL202, and the vertical direction, orthogonally to the bow stick, by an AnalogDevice ADXL103. The sensors are fixed on a board fixed underneath the bow frog by a carbon clamp, made by Alain Terrier.

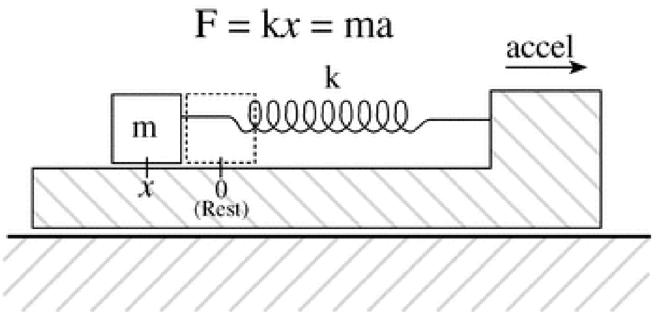


Figure 2.2: Accelerometer mass-spring system. The acceleration of the mass is roughly proportional to its displacement.

The physical principle based behind an accelerometer is that of a mass-spring system 2.2. For each axis, the accelerometer therefore measures the displacements of the mass from its rest position, i.e. spring not stretched nor compressed. The frictions are compensated for by some sophisticated signal conditioning circuitry present in the device. Therefore, the acceleration of the mass is roughly proportional to its displacement.

2.1.3 Measuring the force of the bow on the strings

Another important parameter in the bowed string playing is the pressure of the bow hair on the strings. Directly measuring the downward force on the strings is an issue when trying to build an unintrusive system: indeed, it would imply the insertion of a sensing device between the bow and the string, which would somewhat alterate the violinist play. Alternative solutions that give an order of magnitude of this parameter must be found.

Diana Young implemented a system already used by Anders Askenfelt [2] for the same goal in 1989, based on strain gauges permanently fixed onto the stick. This system can measure the flexion of the bow stick, which is reflecting the normal force between bow and string. It however presented many implementation issues [24]. The carbon fiber heat conductivity is not high enough to dissipate the heat produced by the alimentation of the gauges, which results in unstable measures. The bow is cylindrical and therefore special care had to be taken to adhere the gauges in order to record every deformations of the material. Finally, this system has to be permanently fixed onto the bow, which is opposite to one of our guiding lines: having a sensing system that can be added and removed easily.

Emmanuel Flety chose to add a force sensing resistor (FSR) on the bow to measure the downward force of the forefinger onto the stick. This solution had already been implemented in the HyperCello project.

The sensors are fed by two lithium batteries that are clipped underneath the electronics board. This way the bow is totally wireless.

2.2 Overall Architecture System

The collected data recorded by the sensing system needs to be digitized and sent to a workstation. The accelerometers and the FSR data are sent to a sensor acquisition system via a radio frequency (RF) transmitter (see figure 2.3), while the position data received by the antenna behind the violin bridge is sent via a cable: that way, the augmented bow remains wireless, which is of most importance in order to preserve the violinist way of playing.

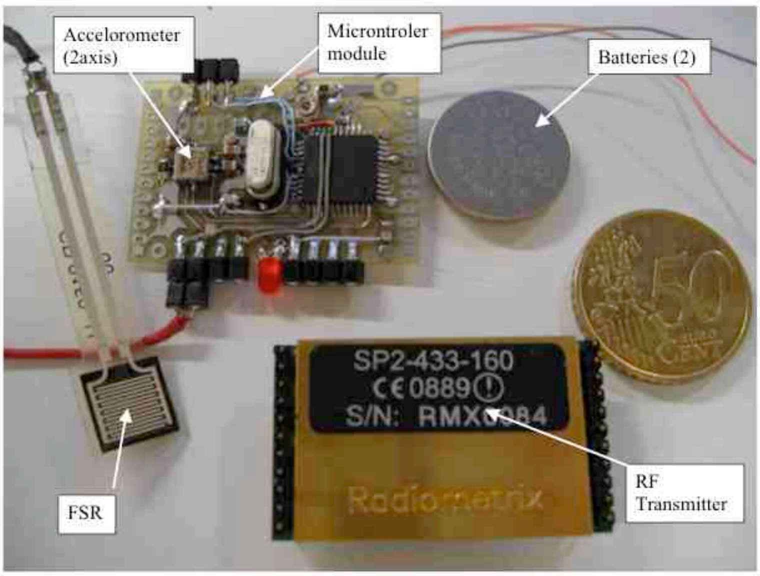


Figure 2.3: The electronics components mounted on the carbon fiber bow.

The digitization device, *EtherSense*, has been developed by Emmanuel Flety and al [3]. *EtherSense* enables the digitization to be performed on 16 bits at 250Hz for the RF transmitter data and at 1000HZ for the position, while in Diana Young system [24], the data are digitized on 8 bits at 41Hz for acceleration and strain and at 142 Hz for position.

The data is then sent via ethernet to Max/MSP using the OSC protocol, where we can record simultaneously audio and gesture data. In a second time, we exported the data to textfiles in order to study them in Matlab in an offline process.

Chapter 3

From violin techniques to physics

This report focuses on the study of several violin bow strokes in order to characterise the sensor signals and bring a simple model that already helps us representing them in a space presenting perceptual coherence. But first, it is necessary to define what a bow stroke is, and qualitatively describe the different variations that can be done on a violin.

3.1 Bow stroke description

It is possible to produce many different sounds with a violin by means of diverse gestures. These modes of play include bow strokes and other techniques where the sound does not result from the action of the bow on the strings, like e.g., plucking the strings with the right or left hand fingers (pizzicato) or bowing the violin body.

A bow stroke can be defined as an articulation of the bow in a single direction to play a note or group of notes. A wide variety of sounds can be produced when bowing the strings in different manners. According to the bow speed, the downward force bow on strings and the elastic properties of the bow hair, it is possible to produce very different timbres.

A qualitative description of some common bow strokes is presented in table 3.1 in order to give an account on their diversity.

In contemporary music, many pieces try to transform the instrument favorite means of excitation. This can be seen with bowed string instruments with a more extensive use of *pizzicato* and *pizzicato bartok* (where the string hits the fingerboard), *con legno* (the violinist plays with the wood of the bow) and *battuto*, where the sound becomes closer to a percussive sound, and *sul ponticello*, *sul tasto* and *ecrasé* which alterate the spectrum distribution by cancelling and emphasizing partials.

Table 3.1: *Description ot some common bow strokes*

bow stroke	**Description**
Détaché	Each note is played in a separate stroke, with a rather constant speed alength the bow and with a more or less smooth attack
Martelé	Strong acceleration at the beginning of each strokes with an abrupt stop between them, which give the stroke a sharp almost percussive attack.
Piqué	Each note is preceded by a pressure on the bow, an accent is given to the note, the attack is sharp but smoother than for Martelé. The bow may leave the string.
Spiccato	Each note is attacked from above the string. The bow describes a sort of parabole and strokes the string when arriving at the parabole lowest points. The violinist plays spiccato around the bow equilibrium point, i.e. around the first third from the frog.
Staccato	Succession of accented notes in one bow, with the bow stopping briefly between two notes. Flying staccato implies the bow leaving the strings and is usually used in arpeggios.
Ricochet	The bow is released from above the strings and bounces with decreasing intensity and time intervals according to its physical properties.
Jetato	The bow is released from above the strings but bounces with constant time intervals so as to achieve a given rhythm.
Saltelato	The violininst performs a short detaché bowing around the equilibrium point, i.e. around the first third from the frog, and the bow naturally bounces under speed. Requires a minimum speed so that the bow can bounce.
Tremolo	Each note is divided in several very short bow strokes played détaché, without the bow bouncing off the string.

3.2 Bow stroke variability and invariance issue

Describing the different bow strokes is not an easy task. Indeed, there are different schools of violin technique, and therefore different ways of performing a bow stroke. The schools do not agree on the terms to use and therefore on the bow stroke classifications. For example, piqué and spiccato may be considered as a same bow stroke. On top of that, a same violinist can perform a bow stroke with many sublte differences for some expressive reasons.

Nuance plays a significant role in this variability. Indeed, a quarter note detache *forte* will be played with a higher bow speed and therefore with more bow length than the same note *piano*, which is of first importance considering our sensing system. In the same manner, the tempo influences the gesture to perform for a given bowstoke. More generally, we can say that the context of a musical phrase can significantly modify a bow stroke, modification totally controlled by expert players.

However, without the data of the musical context, where can be the delimitation between *martelé piano* and *piqué forte*? There seems to be no clear frontier between the different strokes. Therefore, our representation space should reflect the possibility to continuously go from one bow stroke to another, e.g. from détaché to martelé or spiccato.

The notion of invariance is subjacent to this variability. We therefore would be interested in extracting the essential information relative to the execution of a bow stroke, which would be common to different violinists. To what extent can this be done, considering that each violinist has a different body constitution and has adapted his/her technique to it. Violin acoustics, which study the interactions between the bow and the strings and the consequences on the sound spectrum, may bring some elements of response to this invariance issue.

3.3 The acoustics of the violin

The most accepted model of the bowed string is the Helmholtz kink motion. When a bowed string oscillates in steady state, the string can be modeled by two straight lines connected by a kink that rotates in a parabolic path. The interaction between the bow and the string switches between two states: stick and slip. The sticking phase corresponds to the interval when the kink moves between the bow and the nut and during which the string is stuck to the bow hair and therefore takes the speed of the bow. The slipping phase corresponds to the interval when the kink moves between the bow and the bridge and during which the string moves in opposite direction to that of the bow.

In 1973, J.C Schelleng [12] established a diagram representing the region where a Helmholtz motion can be maintained in function of the bow-bridge distance and to the downward force on the string. In addition, Schelleng identified other regions where the bowed string waveform shows spectral properties and therefore associated perceptual adjectives such as "raucous" or "higher modes". This is one of the first works done on the prediction of sound properties from physical parameters.

In [1] and [2], Anders Askenfelt measured various parameters related to the bowing of a violin. He studied the variations of these parameters with different nuances and bowing techniques and showed that the parameters that have a direct impact on sound dynamic level are bow speed, bow-bridge distance and bow force. Later, K. Guettler, E. Schoonderwaldt and A. Askenfelt have shown the influence of bow speed and of bow bridge distance [15], as well as that of bow tilting [6] on the spectrum higher partials.

The studies carried out in acoustics show that the parameters that should be considered in the playing of a violin are the bow speed, the distance between the bow and the bridge and the bow force on the strings. These parameters are strongly correlated to sound characteristics. Therefore, we may expect that focusing on these parameters brings considerable information on the invariance and variability of bow strokes. Ircam's sensing system was built in order to measure these parameters. Before going into a high level interpretation of the sensor signals, we must analyze them at a lower level.

Chapter 4

Low level description

Before interpreting the sensor data to extract invariant patterns, it is important to consider the sensor properties in order to quantify the accuracy of the whole sensing system.

4.1 Discussion on the sensors

We have seen from the studies on acoustics that the parameters of interest include bow speed, bow-bridge distance and bow force on strings. The sensing system should therefore measure these quantities. However, the sensors are not perfect as they may not directly give access to the desired parameter, add noise and have a definite resolution. This part gives a quantification of the sensor performances in the context of the augmented violin bow.

4.1.1 Acceleration sensor signal

In chapter 2, we have seen that the physical principle ruling an accelerometer is that of a mass-spring system (one for each axis): the sensor measures the displacement of a mass connected to a spring in a specified direction. In consequence, the data given by an accelerometer is not a true measure of the system changes of speed. Indeed, it also measures the angle made by the accelerometer axis and the gravity direction, often refered to as *static acceleration*: this term is debatable but as it is commonly used in accelerometer technical specifications, we will used it. The variations of speed are called *dynamic acceleration*. A high-pass / low-pass filtering cannot separate them in all cases, as a violin player can change string during a bow stroke or perform several bow strokes on a single string. In the first case, static acceleration evolves faster than dynamic acceleration while in the second it is the contrary. This double measure is a direct consequence of the physical properties of the accelerometer, and is problematic as the data given by this sensor is the combination of two unknowns: measuring the acceleration using an accelerometer is therefore an under-determined problem.

4.1.2 Position sensor implementation

The main issue here is the implementation of the position sensor. We need to build a homogenous resistance of the bow stick size (60cmx5mm), with a sufficiently low impedance so that the tip and frog signals can gradually decrease alength the bow, and sufficiently high so that the bow does not radiate. The magnetic ribbon made from a video tape still needs some adjustments in order to satisfy this double constraint.

In practice, the electric signals do not decrease linearly along the bow length, as would be wanted to compute the tip to frog distance and the bow bridge distance using the equations given in chapter 2, but have a strong exponential decrease so that each signal amplitude is no more significant beyond the middle of the bow. Therefore, both bow positions are problematic to compute for the moment, because we either get an electric signal from only one extremity (when next to the tip or the frog) or we get both electrical signals with low and noisy amplitude (around the middle of the bow, for about 10 cm).

4.1.3 Force sensing resistor relevance

We chose to measure the force exerted by the forefinger on the bow stick in place of the force bow on strings by means of a force sensing resistor. The weakness of this system is that the FSR signal is not highly correlated to the force bow on strings: one can exert a downward force by explicitly using the forefinger, but at the same time, one could exert a downward force *without* the help of the forefinger but using the other fingers and the weight of the hand. Moreover, this parameter varies significantly, even between violin players of same skills, according to the technique they developed from their body constraints and their instruments.

4.2 Noise estimation

We now examine the noise introduced by the sensing system. We have underlined the technical issues of the various sensors, especially the position sensor. Consequently, we focus on the accelerometer signal in the bowing direction (x-axis), which corresponds to the variations of the bow speed. Table 4.1 shows that the signal noise ratio (SNR) for this signal is correlated to the acceleration values: the higher the values, the higher the SNR.

Since we mainly analyzed the accelerometer signal in the bowing direction in the following parts, we didnot estimate the SNR values for the other sensors.

$Mean$	$\sqrt{Var}$	SNR_{dB}	*Error (%)*	*Comments*
11683	33.8	25.3	0,29%	vertical bow, tip up
14879	29.6	27.0	0,19%	horizontal bow
18304	28.9	28.0	0,16%	vertical bow, tip down

Table 4.1: *SNR values for the x-axis accelerometer. SNR is computed as* $10{*}log(Mean/\sqrt{Var})$*. We can see that noise is correlated to the acceleration values, as the higher the values, the higher the SNR.*

4.3 Range and resolution

4.3.1 Static Acceleration

In table 4.2, we have computed the effective bit resolution for static acceleration, i.e angle of inclination between the bow stick and gravity, according to the signal range and noise. The error value corresponds to the maximum error measurable with the bow being still. As a matter of fact, the resolution is under-estimated.

Min	Max	*Range*	*Error value*	*Resolution*	*Minimum bits needed*
11500	18300	6800	33	206	8 bits

Table 4.2: *Range and resolution for static acceleration in the accelerometer bowing direction*

4.3.2 Dynamic Acceleration

The range of dynamic acceleration has been estimated according to the likely maximum accelerations in a playing situation, which are lower than the accelerometer maximum ones (± 2g). Moreover, the error due to noise has been computed on static measures, i.e with the bow being still. We have seen the SNR being correlated with acceleration. Since, they are computed with lower acceleration values than in a playing situation, we may assume that the error values are over-estimated, and therefore the resolution under-estimated. Table 4.3 shows the results.

Min	Max	*Range*	*Error value*	*Resolution*	*Minimum bits needed*
0	30000	30000	29 to 33	900 to 1035	10 to 11 bits

Table 4.3: *Dynamic range and resolution for dynamic acceleration in the bowing direction*

4.3.3 Velocity computation

Acoustics has shown that there is a relationship between bow speed and sound spectrum, which makes it a parameter to consider in the study of bow stroke variety. There are two ways to compute this speed: from the position signal or from the accelerometer signal. Deriving the position is the most natural way of obtaining the bow speed. However, we have seen the technical issues in building an accurate position sensor and the need for some more adjustments. Integrating the acceleration poses the problem of the unknown initial speed. The other problem is the accelerometer signal being the combination of both static and dynamic accelerations.

In spite of these different problems, it is still possible to extract some information on the bow speed. Even if the tip and frog signals do not decrease as would be expected, we can still derive them. However, the accuracy of this operation remains to quantify. The integration of the acceleration signal with a zero offset can give the global shape of the speed waveform. However, the integration of numeric signals is done by summing all samples, which can result in an error accumulation. In addition, this technique is also limited by the violinist not changing static acceleration during a bow stroke: a change implies a shift of the data that does not correspond to a speed change.

Chapter 5

Violin bow strokes characterization

This report focuses on the violin bow strokes in order to study the wide range of diversity of tones and nuances possible in a bowed string instrument. The study is performed on the basis of the data acquired from the sensing system described in chapter 2. The idea here is to propose a higher level interpretation of the signals and extract invariants and patterns, which results in the creation of a set of features. We will see that the space generated by these features shows pertinent properties in terms of gesture variations and therefore in terms of sound variations.

We have chosen four different bow strokes for this study: *détaché*, *martelé*, *piqué* and *spiccato* (see table 3.1 for a qualitative description of these bow strokes). The recordings include synchronized audio and gesture data.

Acoustics studies suggest three parameters having an influence on sound characteristics. We chose to focus on bow speed, which seems to be carrying the most information for our interest.

5.1 Signal Models

5.1.1 Acceleration

We have focused our analysis on the accelerometer signal corresponding to the bowing direction. Indeed, it represents the variations of the bow speed, which is a parameter having a direct impact on sound according to acoustic studies.

The accelerometer signal is the combination of static and dynamic acceleration and we are interested in the variation of speed. We therefore have to extract dynamic acceleration from the signal. To do so, we recorded a series of each bow strokes performed on a single string. That way, static acceleration remains relatively constant, can be estimated from independent static measures (bow on each strings) and can be substracted as an offset coefficient. Figure 5.1 shows the accelerometer signals for the bow strokes *détaché* 5.1(a), *martelé* 5.1(b), *piqué* 5.1(c) and *spiccato* 5.1(d), performed at the same tempo and nuance, after substraction of static acceleration.

The acceleration signal shows some repeatable patterns. The positive peak at the beginning of the curve represents the quantity of acceleration needed to bring the bow speed from $-speed_b$ to $+speed_b$, which occurs when changing from upbow to downbow. In addition, we can see that all the bow stroke accelerations have a biphasic tendency between positive and negative, which shows that acceleration is followed by deceleration. This amount of deceleration, its time repartition and the moment it occurs is also characteristic to the type of bow strokes. Actually, this makes sense from a physical point of view if we consider that the bow has to be done in a definite bow length: a strong acceleration will necessarily be followed by a certain amount of deceleration in correspondence with the amount of acceleration.

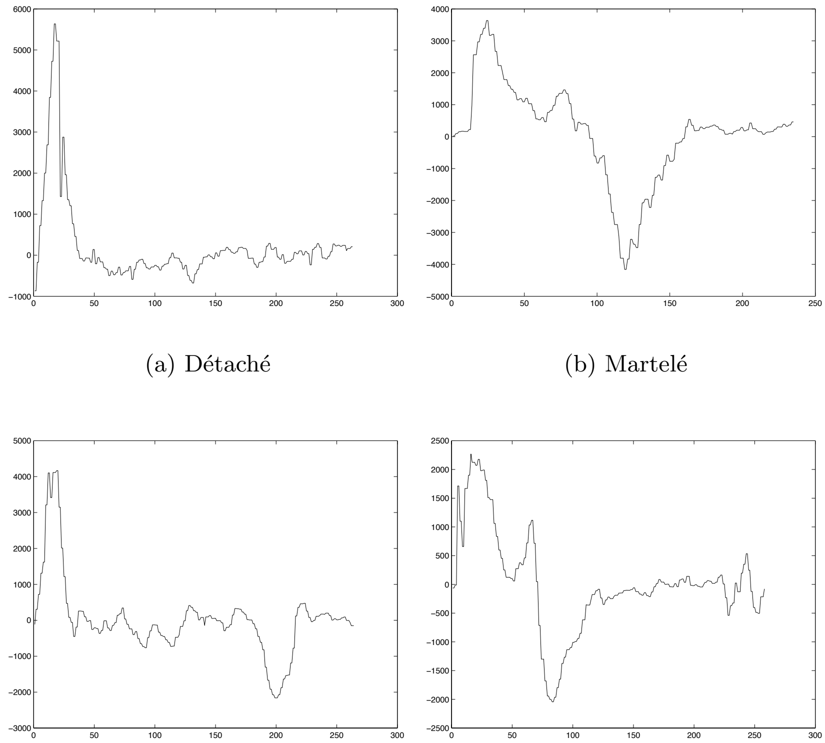
5.1.2 Integrated speed and position

The differences and similarities observed in the dynamic acceleration signals indicate that similar results are to be expected when considering the bow speed and position temporal curves. We integrated the acceleration raw signals with no prior filtering as shown in figure 5.2. The integration of the numeric signals is done by summing the samples over time.

Indeed, bow speed also shows some repeatable patterns for each type of bow strokes. With détaché and piqué, the speed is rather constant while with martelé and spiccato the bow goes faster at the beginning of the stroke before slowing down for the rest of time. The considered bow strokes particularly show differences in their attack (first 200ms) and their final speed.

5.1.3 Audio signal correlation

Acoustic studies stressed the strong influence of bow speed on dynamic sound level. It is interesting to discuss to what extent we can find the gestural data in the audio. Indeed, when keeping bow-bridge distance constant and with no vibrato, we can see that the bow speed resulting from the integration of the dynamic acceleration signal and the audio energy are strongly correlated, as shown on figures 5.3(a) and 5.3(b). However, in a playing situation bow-bridge distance and bow speed varies simultaneously, and the tones are played vibrato. Audio signal results from all the violinist gestures. Audio cannot separate both information while a violinist can play indepedently on each parameter. We may therefore infer that studying the audio in order to characterize the subtle differences in the violin playing will not be as fine an analysis as directly studying gestural information, because all



(a) Détaché (b) Martelé

(c) Piqué (d) Spiccato

Figure 5.1: Accelerometer signals in the bowing direction after substraction of static acceleration. The signals are biphasic between positive and negative values. Some differences can be observed in the amount of deceleration, its time repartition and its occuring moment. The abscissae represent the sample number with sampling frequency 250Hz.

is packed in one signal: the vibrato is an alteration of the pitch, while bow speed and bow-bridge distance variations are relative to the production of sound.

5.2 Segmentation

5.2.1 Segmentation objectives

The sensing system records the violin player bowing gestures. Therefore, we receive a set of time series corresponding to each sensors and from which we try to extract information. In the recordings we made, each bow stroke is played repeatedly. The very first task that needs to be done in order to characterize the bow strokes is to segment the time series according to each bow strokes, i.e. part downbows and upbows. An automatic segmentation algorithm would help us constitute a large bow stroke database and therefore have a wider range of variations for each bow stroke.

5.2.2 Automatic segmentation issue

The most straighforward idea is to automatically segment the data according to the speed signal. Indeed, whatever the bow stroke, the bow speed shifts from $+speed_b$ to $-speed_b$ between downbow and upbow. As there is no speed sensor on the system, it has to be computed either from the position sensor signal, or from the accelerometer signal. As we underlined, both approaches have their difficulties. We also discussed the position sensor implementation issue, i.e. the building of a resistance of the bow size with the right impedance. Thus, the first analysis we performed were done without the position sensor, which needed improvements. In consequence, we focused on the accelerometer signal in the bowing direction, and implemented a segmentation algorithm based on thresholds and peak detection.

The accelerometer signal is the combination of static and dynamic accelerations. This particularity is a problem if we want to use a threshold method to segment the data. Indeed, a high-pass / low-pass filtering cannot separate them in all cases, as a violin player can change string during a bow stroke or perform several bow strokes on a single string. In the first case, static acceleration evolves faster than dynamic acceleration while in the second it is the contrary.

Using the acceleration signal, we did not try to implement a threshold algorithm that would have worked in all cases because this would have demanded many heuristics and therefore be little robust. We instead reduced the possible cases by asking the violin player not to change string during the recording of a bow strokes series. This choice did not prevent us from performing measures on all strings but eased the substraction of static acceleration.

5.2.3 Segmentation procedure

We already observed different behaviors on the dynamic acceleration according to the bow strokes. Downbows are have an acceleration peak at the beginning, are biphasic between positive and negative and some show strong deceleration peaks, and inversely for upbows. Therefore, the segmentation algorithm must consider this deceleration peak as a part of downbow and not consider it the beginning of upbow 5.4.

The segmentation algorithm is a two step process. It first thresholds the signal according to a user given value in order to cancel the signal lowest values variations keeping the sharp peaks. Then, by differentiating the time instant array returned by the thresholding process, we can find the bow stroke change instants with another threshold value related to the minimum interval possible between two peaks: this value depends on tempo and therefore must be chosen of the same order of magnitude.

The bow strokes are all performed on a single string. We chose to estimate static acceleration as the mean of the accelerometer output over a window containing 10 to 100 bow strokes. This estimation is convenient because the same blind treatment can be applied to all bow strokes. Figure 5.5 shows the times series for bow strokes performed at 60bpm. Static acceleration has been removed by substraction of the mean value over the whole window. The first column corresponds to the raw dynamic acceleration, the second column is the filtered, thresholded acceleration with a hann window of size 64 (which corresponds to 250ms) and the last column is the filtered, thresholded acceleration with the markers at each bow change.

Because of the détaché, martelé, piqué and spiccato differences in the dynamic acceleration signals, careful adjustments had to be done on the threshold values for each of them. However, once the appropriate range of thresholds is found, the algorithm can segment the data at different nuances and tempi, and is very useful to build a larger bowtroke database than with a purely manual segmentation.

The main objective of this segmentation algorithm is to help us consitute bigger databases in order to study the bow strokes invariance and variability. However, segmenting the bow strokes is already a relevant operation considering the different applications. Indeed, a real-time version of a robust segmentation algorithm can be used to track the interpret movements in a score following like application, or can be used to trigger events in a mixed piece. Having the bow speed zero crossing may be of great help to achieve segmentation in real time.

5.3 Features

From the signal modelisation of the bow strokes, we extracted six simple features, relative to the bow speed and its variations. Considering the dynamic acceleration on a window corresponding to the stroke duration, we compute:

- the maximum acceleration, a_{max}
- the minimum acceleration, a_{min}
- the time of maximum acceleration, t_{max}
- the time of minimum acceleration, t_{min}
- the speed after maximum acceleration : maximum speed, v_1

- the speed after minimum acceleration, v_2

These features are simple. Considering the speed evolution curve during a bow stroke, they correspond to the most common analytic parameters, in a mathematical sense, that can be extracted (extrema, inflexion points and time interval) (fig 5.6). They impose some geometric constraints on the speed temporal curve.

From a violinist and physical point of view, the features can be interpreted as follow:

- v_1 is the speed of attack

- v_2 is the speed after the attack

- a_{min} and a_{max} are relative to the sharpness of the attack

- t_{min} and t_{max} are relative to the time of the attack

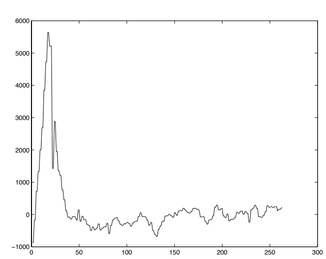
Another point is that these features are all extracted from the shape of the bow speed. We have discussed the correlation of these parameters with the audio signal. These features therefore appear to be relevant in our interest in the subtle timbres and nuances variations that can be produced on a violin. We can also notice that since they have a physical interpretation, they conceptually make sense in terms of playing technique from a violin player point of view. We combined these features in order to consider relative values instead of absolute values. The set of combined features is therefore:

- the maximum speed, v_1

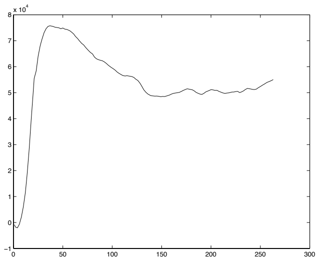
- the normed speed, $v_2 * |\frac{a_{min}}{a_{max}}|$

- the time difference, $\Delta t = |t_{max} - t_{min}|$

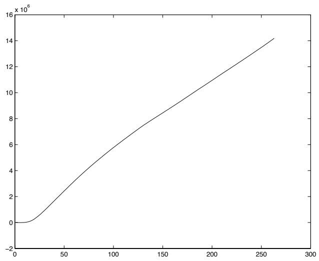
This combination directly reflects the geometrical constraints of the speed curves: especially $v_2 * |\frac{a_{min}}{a_{max}}|$ highlights the relations between the acceleration area (or speed), the acceleration peaks and their time repartition. It is to be noted that this combination may not be optimum in terms of performances but already gives good and interpretable results. It is also to be noted that given that we consider $\Delta t = |t_{max} - t_{min}|$, we are invariant to the probable artefacts brought by the segmentation algorithm.



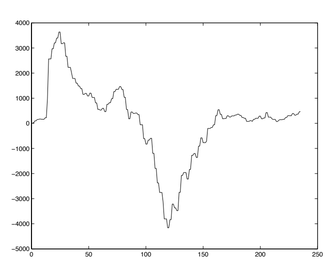
(a) Détaché acceleration



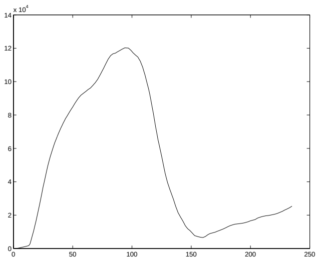
(b) Détaché integrated speed



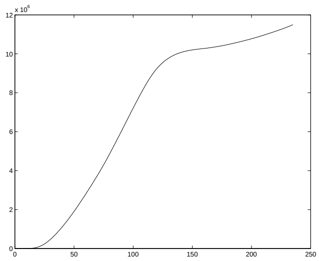
(c) Détaché integrated position



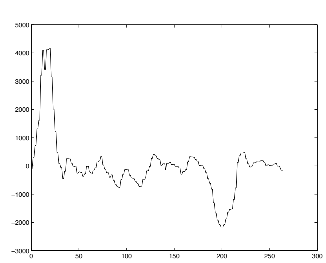
(d) Martelé acceleration



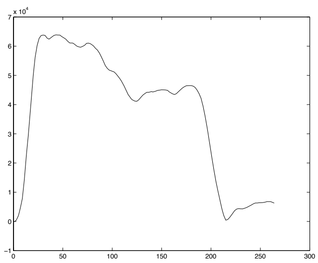
(e) Martelé integrated speed



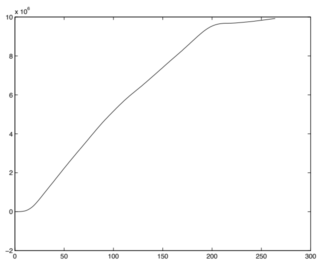
(f) Martelé integrated position



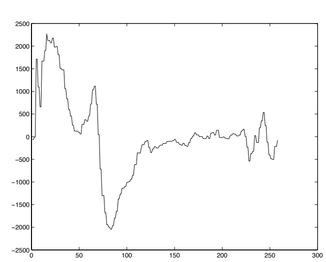
(g) Piqué acceleration



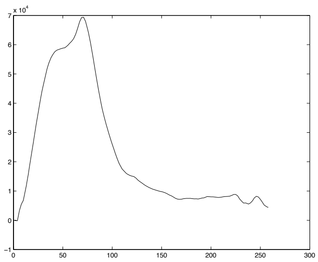
(h) Piqué integrated speed



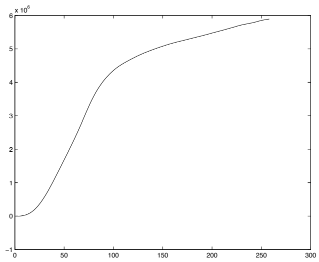
(i) Piqué integrated position



(j) Spiccato acceleration

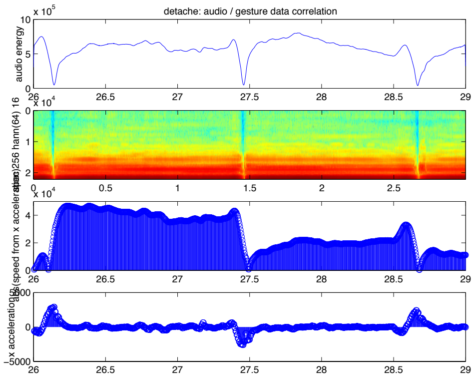


(k) Spiccato integrated speed

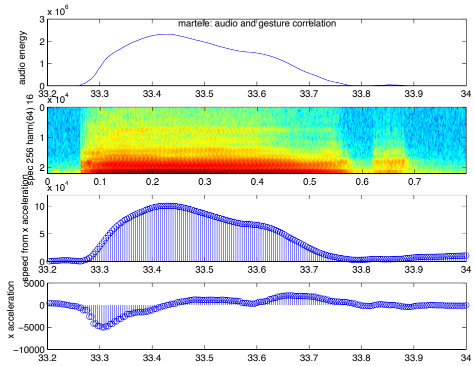


(l) Spiccato integrated position

Figure 5.2: Acceleration, integrated speed and integrated position signals. Speed and Position are integrated by summing the acceleration samples with a zero offset. The abscissae represent the sample number with sampling frequency 250Hz.

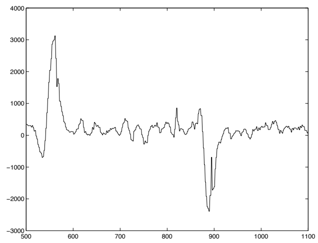


(a) Several détaché bow strokes

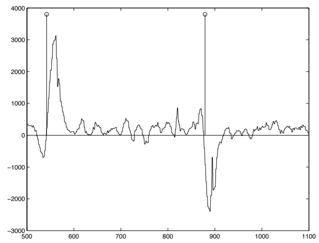


(b) Martelé bow stroke

Figure 5.3: Audio - Gesture data correlation. From top to bottom: audio signal energy, audio spectrogram, integrated speed absolute value, dynamic acceleration. X-axis in seconds

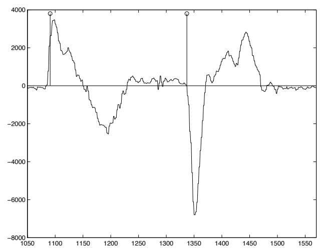


(a) Détaché acceleration



(b) Segmentation markers

(c) Martelé acceleration



(d) Segmentation markers

Figure 5.4: Bow Stroke Segmentation. 5.4(a) and 5.4(b) represent the acceleration signal and the bow stroke segmentation for two détaché bow strokes. 5.4(c) and 5.4(d) represent the acceleration signal and the bow stroke segmentation for two martelé bow strokes. There is a strong deceleration peak in martelé which is part of the execution of the bow stroke and is not the beginning of the following one.

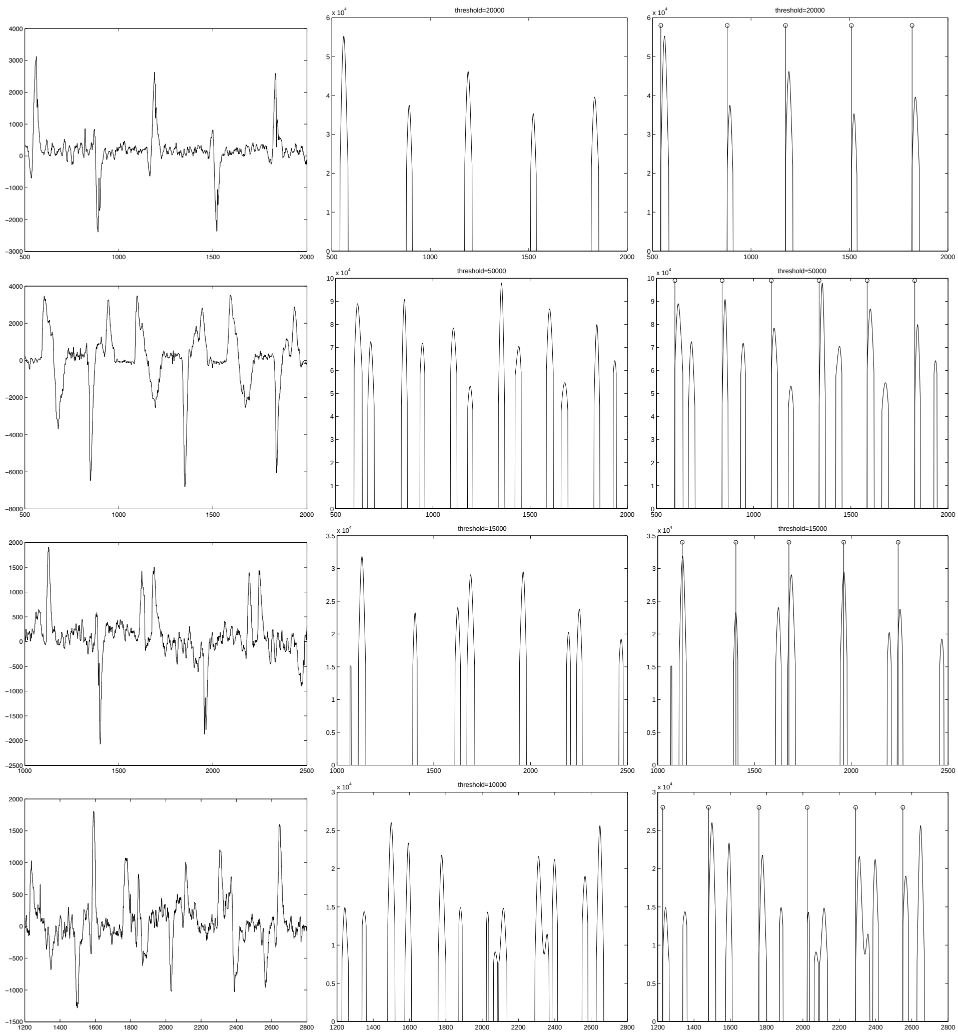


Figure 5.5: Segmentation Steps. Each line concerns a different bow stroke: from top to bottom, détaché, martelé, piqué and spiccato. From left to right: raw dynamic acceleration, filtered and thresholded signal with a 64-hann window and the segmentation markers.

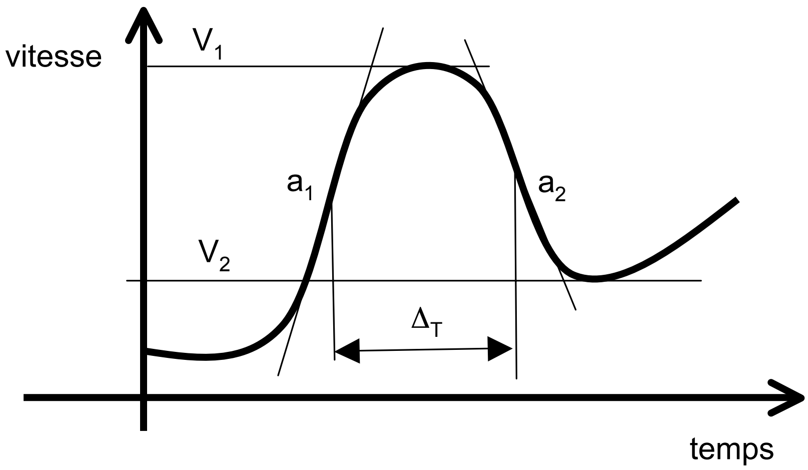


Figure 5.6: Speed features. On this graph, $a_1 = a_{max}$, $a_2 = a_{min}$.

Chapter 6

Analysis of a bow stroke database

We have tested our set of features on a bow stroke gesture database elaborated thanks to our segmentation algorithm. The gesture data comes from a violin player using the augmented bow and a violin we provided. We examine the coherence of the set of features in terms of gesture invariance and variations.

6.1 Measurement Protocol

The violin player was asked to perform some series of bow strokes for each type détaché, martelé, piqué and spiccato. Each type of bow strokes was performed on different strings (one for each measure), with different left-hand fingering, nuance and tempo.

We simultaneously recorded audio by a cardioïd KM-140 microphone placed at about 50cm to the violin. The audio was digitized by a MOTO 828 sound card. Synchronization with gestural data was done by Max/MSP. In order to align sound and data, we triggered a Heaviside function coupled with a sinusoid wave and recorded the former as gestural data and the latter as sound.

The data was exported to textfiles in order to study it in Matlab. The segmentation algorithm described in chapter 5 was then used to constitute the bow stroke database. We extracted the features described in chapter 5 from each of the database bow strokes and plotted them in the space

$$\{v_1; v_2 * |\frac{a_{min}}{a_{max}}|; \Delta t\}$$

6.2 Results

The features are plotted in the figure 6.1. Each point represents a bow stroke played in a certain way: blue color is for détaché, red for martelé, green for piqué and black for spiccato. The plotted data represents the feature values for notes played moderato (60 bpm), *mezzo forte*, on each of the four strings. The top plot on the left is a 3D view of the bow strokes feature points. This 3D space is generated by v_1, $v_2 * |\frac{a_{min}}{a_{max}}|$, and Δt. The three other plots are the projection of the data on each coordinate plane.

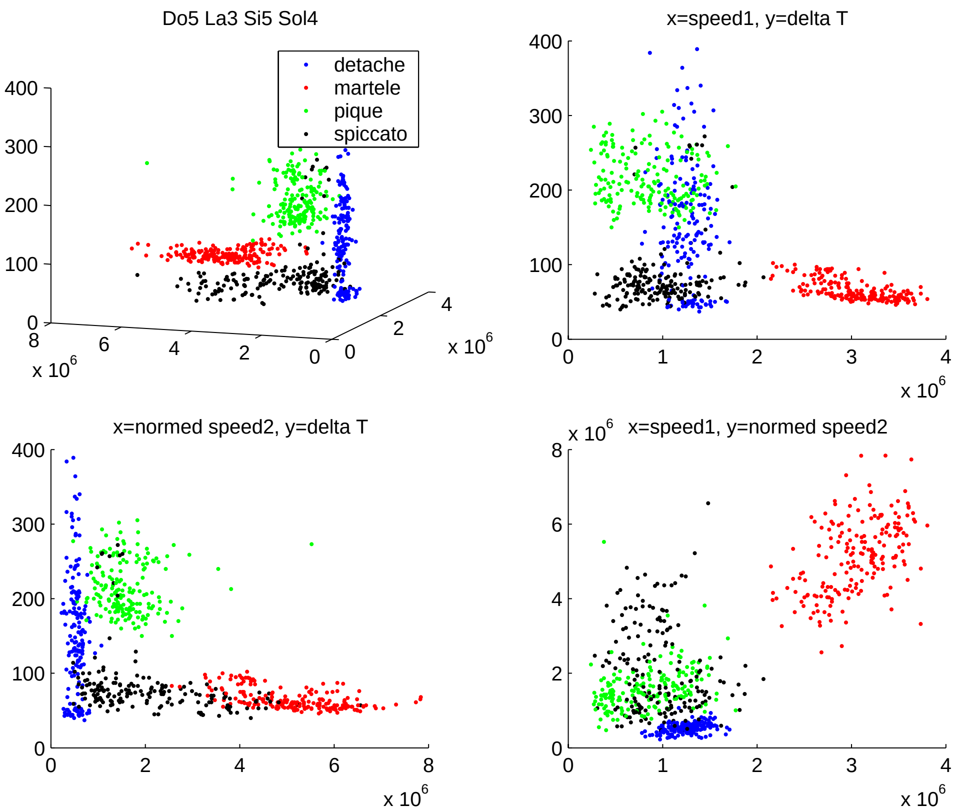


Figure 6.1: Features of bow strokes for different tones played moderato (60 bpm), *mezzo forte*, on the four strings. Top left plot is a 3D representation of the clusters. The three other plots are projections on the coordinate planes. Each point represents a bow stroke played in a certain way. Blue is for détaché, red for martelé, green for piqué and black for spiccato. There are approximately 200 points per type of bow strokes.

We can identify four distinct clusters corresponding to the four types of bow strokes we analyzed. The features show a first invariance property in that they stay clustered whatever the string played. This property is not obvious considering that playing on the G string at the same nuance as on a E string does not demand the same gestural effort.

It is to be noted that the détaché bow stroke cluster repartition is mainly along the Δt axis, which was predictable given that the speed remains roughly constant over the execution of the bow stroke: this dimension is therefore not meaningful for this bow stroke. Martelé and spiccato have a rather constant Δt value, which indicates how characteristic this feature is for those bow strokes. Moreover, if we do not consider the Δt axis, which somewhat corresponds to not waiting the end of the bow stroke, we can see that martelé will still be separated but that the other bow strokes will not be. So, in order to fully characterize the different types of bow strokes we studied, it is necessary to wait for the bow stroke attack to end.

We have discussed the strong correlations between bow speed and audio energy. We can therefore infer that the clusters representing the bow stroke features also have a perceptual pertinence. This fact would be to study more in details in the future with psycho-acoustics studies.

6.3 Gesture variations and feature points behavior

We now examine the feature space properties more deeply considering changes in nuance and tempo.

6.3.1 Nuance variations

Figure 6.2 plots the feature points for a note played moderato (60 bpm) at the nuances *pianissimo* (symbol +), *mezzo forte* (symbol .) and *fortissimo* (symbol x).

The first result is that the points are still clustered, which reinforce our invariance property. Now, if we observe the points more in details, we can see that a modification in nuance results in variations on v_1 and on $v_2 * |\frac{a_{min}}{a_{max}}|$, i.e. variations on speed amplitude rather than on the time interval between the acceleration extrema $|t_{max} - t_{min}|$. This can be explained by the attacks being less marked in softer nuances than in louder nuances. We therefore can determine variation directions directly related to nuance variations: a bow stroke performed louder will result in its feature point having a higher v_1 and on $v_2 * |\frac{a_{min}}{a_{max}}|$, and vice versa.

The feature space clusters show some interesting relevance. Indeed, we can see that the points corresponding to piqué *fortissimo* are close to martelé *pianissimo*, which seems pertinent from a violinist point of view. More generally, *fortissimo* bow strokes tend to be more marked and therefore become closer to martelé, as for spiccato and détaché.

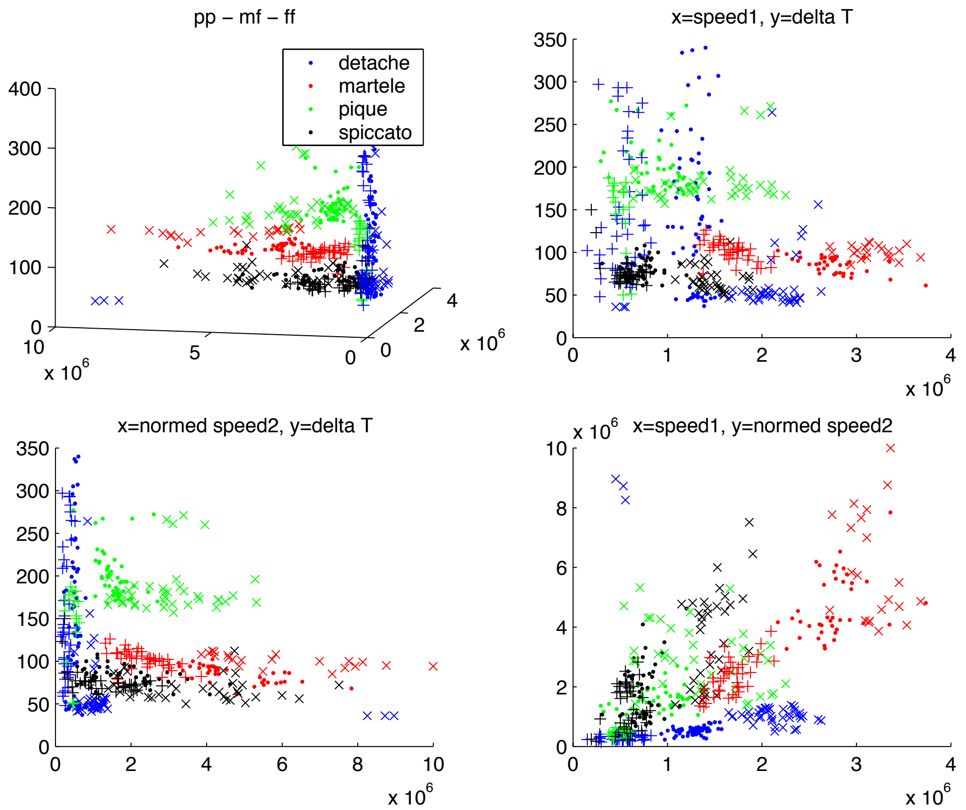


Figure 6.2: Nuance variation. (symbol +) is for *pianissimo*, (symbol .) for *mezzo forte* and (symbol x) for *fortissimo*. For visibility convenience, only 30 points were plotted per nuance and per type of bow stroke. The 3 lost blue crosses result from a bug in the segmentation algorithm

6.3.2 Tempo variations

This time we focus on the feature space behavior with varying tempo. Figure 6.3 plots the feature points for a note played *mezzo forte* with the tempi moderato (60 bpm, symbol .), and allegro (120bpm, symbol Δ).

The clusters are still coherent when varying tempo. However, the Δt value is not constant whatever the tempo but decreases when tempo increases. The variations along this axis remains to examinate.

As with variations of nuances, we can see some pertinent variability in the feature space, specially the getting closer of piqué allegro and martelé, which again seems pertinent from a violinist point of view.

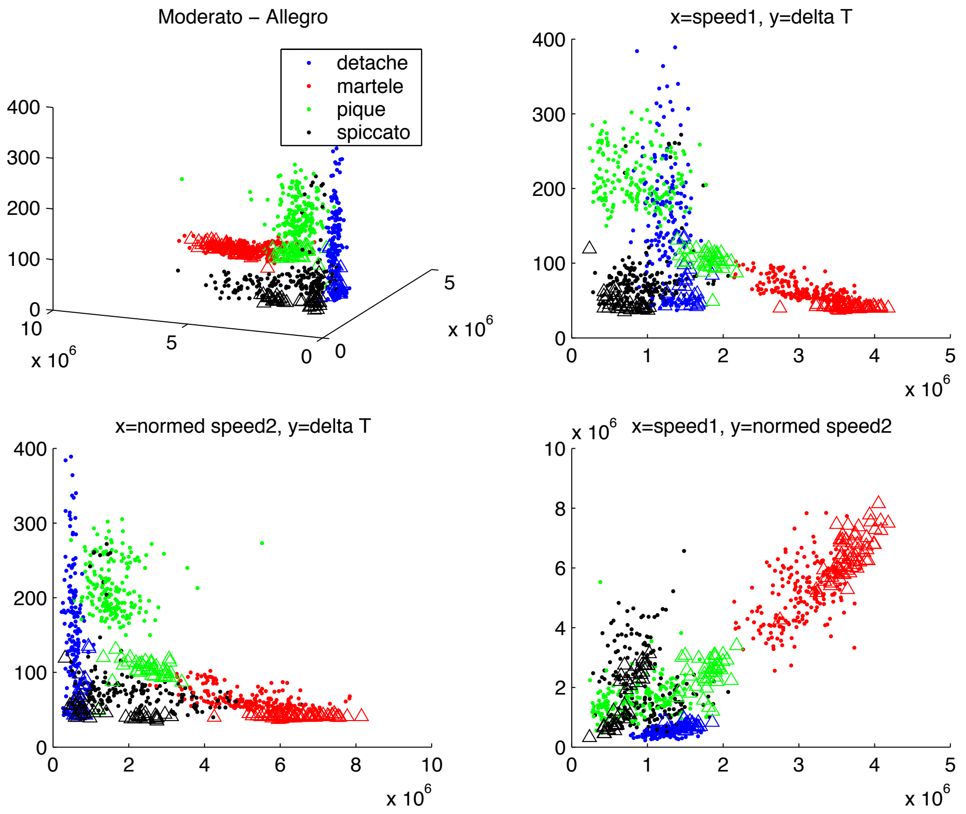


Figure 6.3: Tempo variations. Notes are played *mezzo forte* moderato (60 bpm, symbol .) and allegro (120bpm, symbol Δ).

6.3.3 Further variation characterization: cluster overlapping

Figure 6.4 plots the variations according to tempo and nuances. The feature points correspond to a note played *pianissimo* (symbol +), *mezzo forte* (symbol .) and *fortissimo* (symbol x) at a moderato tempo, and *mezzo forte* allegro (symbol Δ).

This figure illustrates the overlapping of bow stroke feature points. More particularly, *mezzo forte*, allegro piqué (green Δ) and *pianissimo*, moderato martelé (red +) points cluster in the same region. We do not know whether this artifact is due to a problem with our features or if the gestures are exactly the same. There are at least two ways of getting some clues about it. We have seen the strong correlation between bow speed and audio. A detailed spectrum analysis might help answering the question. The other way concerns psycho-acoustic studies. We have indeed stressed the perceptual pertinence of the feature space. It would therefore be interesting to see if subjects reproduce the same confusion.

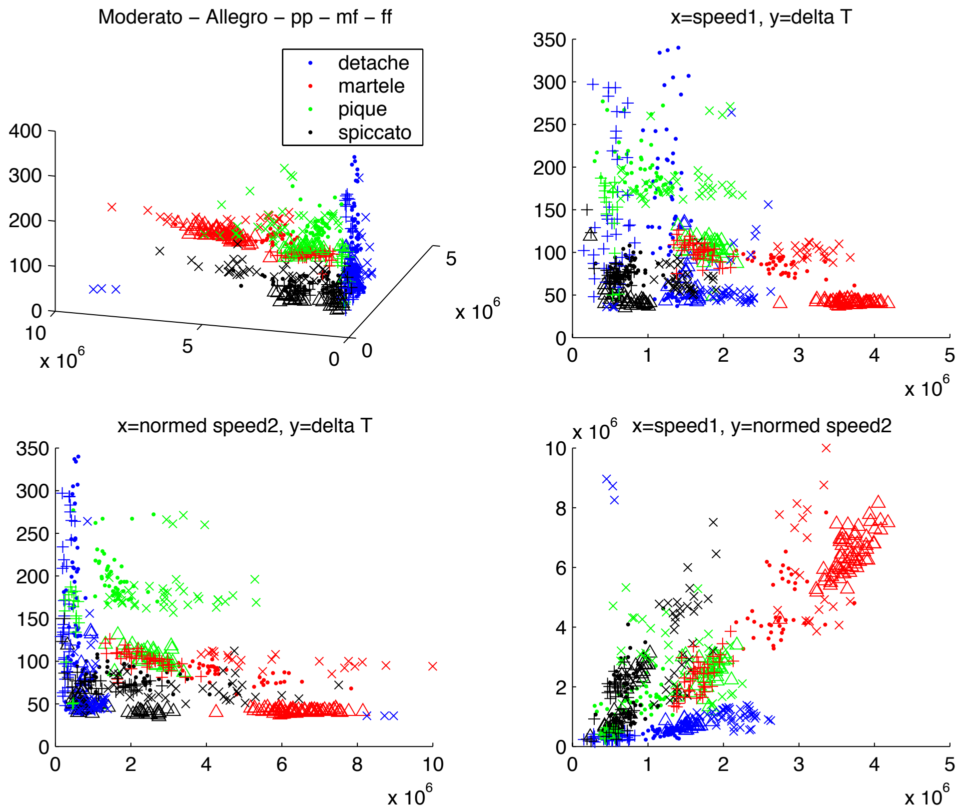


Figure 6.4: Nuance and Tempo variations. Notes are played *pianissimo* (symbol +), *mezzo forte* (symbol .) and *fortissimo* (symbol x) at a moderato tempo, and *mezzo forte* allegro (symbol Δ)

6.3.4 Player variations

We asked professional violinist Jeanne-Marie Conquer to perform the bow-strokes described in the measurement protocol. Figure 6.5 shows the feature points from the analysis of her movements (points marked +). The points marked (.) are from measures with my own movements. It is to be noted that only three types of bow strokes have been performed by Jeanne-Marie Conquer, for whom piqué and spiccato are a same bow stroke.

The clusters remain coherent between the two players, which shows a strong invariance property. However, we should keep in mind that the performed bow strokes may be stereotyped, because out of a musical context. Further analysis should be carried out on musical excerpts, and with more players of different skills, in order to study finer playing differences.

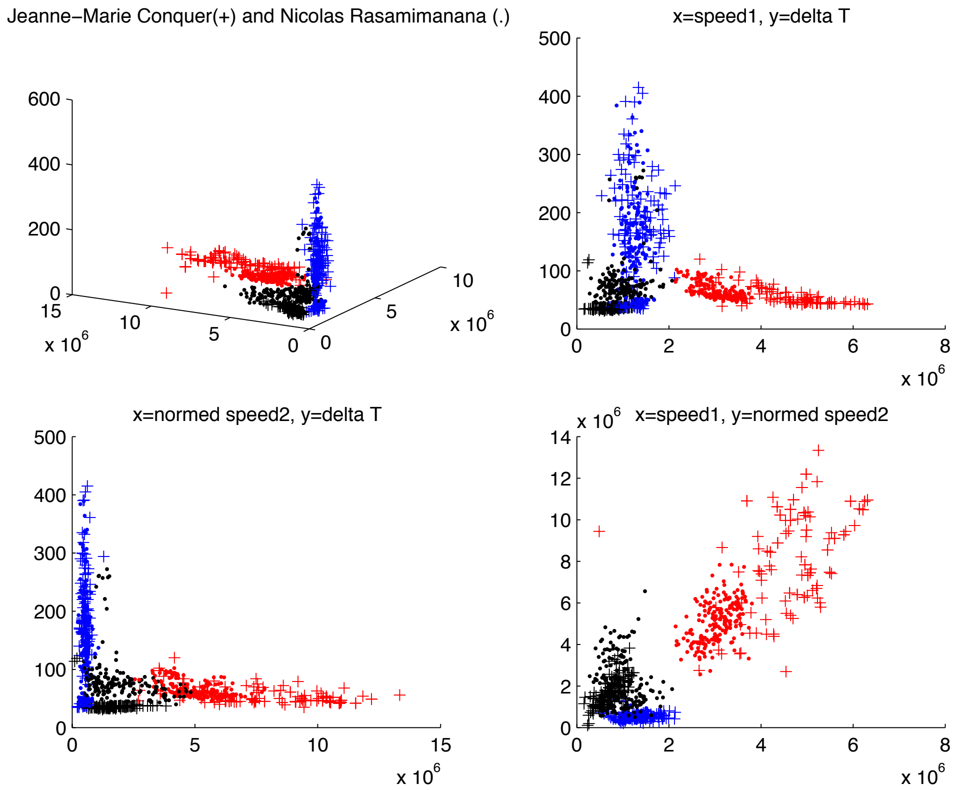


Figure 6.5: Player variations. The feature points marked (+) are relative to violonist Jeanne-Marie Conquer. The features marked (.) are mine. Blue is déraché, Red martelé and Black spiccato. The clusters show a strong invariance property.

6.3.5 Discussion on the features

The set of features we extracted is simple: it would be very surprising if it could represent all the variations possible with a violin bow. However, it already shows some pertinent properties when considering the bow stroke types détaché, martelé, piqué and spiccato and their variations according to nuances and tempi. In order to complement the feature space pertinence, further investigations needs to be carried out with other bow strokes. Moreover, it would also be interesting to study gestures from more violinists. Beyond the comparison of violin players, which would greatly trigger interest to the most radical violin schools, the idea is to see how pertinent this feature space can be for each of the violinists. The results with two players suggest that we need to study bow strokes in a real playing situation, i.e while playing a musical piece, which would reveal more details on each of the violinist expressive possibilities with the bow.

Chapter 7

Conclusion and perspectives

This report presents results on gesture analysis using an augmented violin. We have built a set of features derived from the violin bow speed, which is a parameter having a strong correlation with audio spectrum according to acoustic studies. We constituted a bow stroke database and analyzed it using our segmentation and feature extraction algorithms. The feature space shows some invariance properties, as the four bow strokes we analyzed are represented in four distinct clusters. It also shows some pertinent behavior according to variations in tempo and nuance. The set of features being derived from bow speed, and bow speed being strongly related to audio spectrum imply that the feature space also has a perceptual pertinence.

The study we performed in this report has been carried out in an offline process. The musical applications evoked in the introduction of this report implies a real time version of this analysis. The segmentation of the bow strokes can possibly be performed in real time with an accurate position sensing system by scanning the signal frame by frame and observing the position derivate zero crossing. Concerning the bow strokes feature computation, an ontological problem arise: we cannot tell what bow stroke is being played before it is finished. Therefore, if we want to use a complete bow stroke characterization, only soft real time can be done, i.e. a response with a constant delay. However, a partial characterization of the bow stroke being performed already brings a lot of information and can be compliant with real-time. We next may use our set of features as a base in a probabilistic approach in the bow stroke temporal development. The most relevant method in this case seems to be Hidden Markov Models.

The results of gesture analysis can be exploited in the score following scheme. A real time segmentation of the gesture data according to the bow stroke changes can provide useful complementary information to determine where the interpret is in the score.

The report also underlines the relation between the features derived from bow speed and perceptual characteristics. The study of the characteristics of sound and its perception in parallel to gestures is to be deepened. Indeed, in addition to the correlation between bow speed and sound spectrum we highlit, there is a constant feedback between the interpret movements and the sound produced: there is no musical gesture without sound. A psycho-acoustic analysis may bring a more precise idea of the correlation between gestures and perceived sounds. More particularly, a free categorization procedure on our bow stroke database sounds would bring some interesting clues on our feature clusters, especially concerning possible confusions.

It is also to be noted that although the set of features we built is only based on dynamic acceleration in the bowing direction, we already get some promising results on our bow stroke database. Further studies will show the pertinence of other sensor data and examine the relevance of modifying the sensors and/or adding new ones. We are considering placing several FSRs on the bow stick in order to measure the pressure of all the fingers on the bow. This may give a richer information than the forefinger pressure on the bow. In addition, in the chain that goes from motor orders output from the brain to the gesture being performed, the violin bow is the last interface before producing sound. It may therefore be pertinent to examine what information can be extracted from the violin player body. This study would bring some rich information on how the whole interpret body moves to produce a sound.

Bibliography

[1] Anders Askenfelt. Measurement of bow motion and bow force in violin playing. *J. Acoust. Soc. Am.*, 80(4), 1986.

[2] 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.

[3] 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.

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

[5] Camille Goudeseune. *Composing with parameters for synthetic instruments*. PhD thesis, University of Illinois Urbana-Champaign, 2001.

[6] K Guettler, E Schoonderwaldt, and A Askenfelt. Bow speed or bowing position - which one influences spectrum the most? In *Proceedings of the Stockholm Music Acoustics Conference, SMAC*, 2003.

[7] Camille Goudeseune Guy Garnett Timothy Johnson. An interface for real-time classification of articulations produced by violin bowing. In *Proceedings of the Conference on Human Factors in Computing Systems*, 2001.

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

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

[10] Charles Nichols. The vbow: Development of a virtual violin bow haptic human-computer interface. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2002.

[11] 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.

[12] J.C. Schelleng. The bowed string and the player. *J. Acoust. Soc. Am.*, 53(1), 1973.

[13] Bernd Schoner, Chuck Cooper, Chris Douglas, and Neil Gershenfeld. Data-driven modelling and synthesis of acoustical instruments. In *Proceedings of the International Computer Music Conference, ICMC*, 1998.

[14] 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.

[15] E Schoonderwaldt, K Guettler, and A Askenfelt. Effect of the width of the bow hair on the violin string spectrum. In *Proceedings of the Stockholm Music Acoustics Conference, SMAC*, 2003.

[16] 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.

[17] Stefania Serafin and Diana Young. Toward a generalized friction controller: from the bowed string to unusual musical instruments. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.

[18] Atau Tanaka. *Trends in Gestural Control of Music*, chapter Musical Performance Practice on Sensor-based Instruments, page 389 to 406. Ircam, 2000.

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

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

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

[22] Diana Young. New frontiers of expression through real-time dynamics measurement of violin bows. Master thesis, M.I.T, 2001.

[23] Diana Young. The hyperbow: A precision violin interface. In *Proceedings of the International Computer Music Conference, ICMC*, 2002.

[24] 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.

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

[26] Diana Young and Stefania Serafin. Playability evaluation of a virtual bowed string instrument. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2003.