Gesture Analysis of Violin Bow Strokes

Nicolas H. Rasamimanana, Emmanuel Fléty, and Frédéric Bevilacqua

IRCAM, 1, Place Igor Stravinsky, 75004 Paris
{Nicolas.Rasamimanana, Emmanuel.Flety, Frederic.Bevilacqua}@ircam.fr

Abstract. We developed an "augmented violin", i.e. an acoustic instrument with added gesture capture capabilities to control electronic processes. We report here gesture analysis we performed on three different bow strokes, *Détaché*, *Martelé* and *Spiccato*, using this augmented violin. Different features based on velocity and acceleration were considered. A linear discriminant analysis has been performed to estimate a minimum number of pertinent features necessary to model these bow stroke classes. We found that the maximum and minimum accelerations of a given stroke were efficient to parameterize the different bow stroke types, as well as differences in dynamics playing. Recognition rates were estimated using a kNN method with various training sets. We finally discuss that bow stroke recognition allows to relate the gesture data to music notation, while a bow stroke continuous parameterization can be related to continuous sound characteristics.

Keywords: Music, Gesture Analysis, Bow Strokes, Violin, Augmented Instruments.

1 Introduction

There is an increasing interest in using gestural interfaces to control digital audio processes. Despite numerous recent achievements ([12]), important ground work on gesture analysis is still necessary for the improvement of such interfaces. We are currently developing various "augmented instruments", i.e. acoustic instruments with added gesture capture capabilities. Such an approach remains remarkably fruitful for the study of gesture in music. As a matter of fact, the use of acoustic instruments in this context allows to apprehend instrumental gesture in a *a priori* defined framework, linked to both a symbolic level, the music notation, and a signal level, the acoustic instrument sound.

One of our current project concerns an "augmented violin", similar to the one developed by D. Young [13]. On a fundamental level, our goal is to build a model of the player's gestures reflecting his/her expressive intentions related to violin playing techniques. Specifically, our aims are to establish the relationships between the captured data, bowing styles and sound characteristics. This includes the study, on a gestural level, of the variations that occur between different interpretations of a single player or between players. These studies will

S. Gibet, N. Courty, and J.-F. Kamp (Eds.): GW 2005, LNAI 3881, pp. 145–155, 2006.

lead us to the development of real-time analysis tools, enabling an *interpretation feedback*, which includes gesture recognition, and *gesture following*, i.e. the possibility to track a performance with respect to a predefined reference. We believe that both approaches are key to develop novel types of interaction between instrumentalists and computers.

We report in this paper the study of three violin bow strokes (*Détaché*, *Martelé* and *Spiccato*) and the evaluation of their possible recognition. The article is organized as follows. We first present a review of similar works. In section 3, we present the capture system implemented on the violin. In sections 4 and 5, we show results on the parameterization and recognition of bow stroke types. Finally, we conclude in sections 6 and 7 by a discussion of these results and their implications on future work.

2 Related Works

Our concept of "augmented instruments" is similar to the Hyperinstruments developed by T. Machover and collaborators. The idea is to use a traditional instrument and to extend its capabilities by digital means. For example, the HyperCello [7] created in 1991 was conceived as an acoustic cello with added measurements of wrist movement, bow pressure and position, and left hand fingering. More recently, D. Young extended the HyperCello to the violin with the HyperBow [13], [14].

Several other interfaces have been developed based on string instruments for artistic purposes ([6], [4], [11], and [9]). All of these works generally used the sensor signals to directly control sound treatment parameters, such as filters [11] or physical model synthesis [14]. B. Schoner [10] adopted a probabilistic approach to infer, in real time, cello sounds from the gesture input given by the HyperCello.

Very few works actually report an analysis of the signals, specifying the relationships between the data and the instrumentist's performance. Among them, C. Peiper et al. [8] used decision tree techniques to classify violin bow strokes based on motion tracking. We here pursue the approach of analyzing different types of bow strokes, and in particular we propose to estimate invariance and variability of the measured signals.

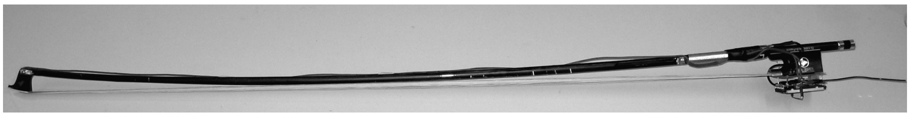
3 Hardware Design

Hardware developments were designed with the following constraints: compatibility with an acoustic violin, no significant alteration of the instrument, wireless communication, relatively inexpensive. The prototype we built and used in this study is shown on figure 1. Two types of gesture data are measured, using technology similar to the one described in [13]: bow position and bow accelerations.

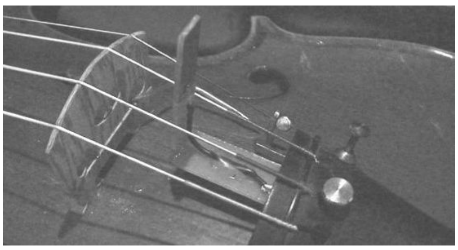
First, the sensing system can measure the bow-strings contact position along two directions: between tip and frog, between bridge and finger-board. This position is measured via capacity coupling between a resistive tape fixed along the bow and an antenna behind the bridge.

Second, acceleration is sensed thanks to two Analog Device ADXL202 placed at the bow frog. Note that such sensors are sensitive to both gravity, hence inclination, and movement acceleration (generally referred as static and dynamic accelerations). The two accelerometers are fixed to the bow nut in such a way that acceleration is measured in three dimensions: bowing direction, string direction and vertical direction.

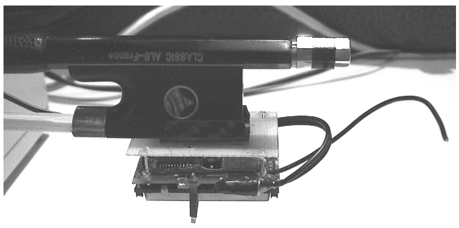
The position data, obtained from the antenna behind the bridge, is digitized in 16 bits with a sensor acquisition system developed at IRCAM, *Ethersense* [3]. The acceleration data are sent wirelessly to a RF receiver also connected to the sensor acquisition system. The acceleration dynamic range has been measured to be of 65 dB. All the data are transmitted to Max/MSP through an ethernet connection using the Open Sound Control protocol, at a data rate of 200 Hz. The surplus weight added by the sensing system is actually 15 grams, mainly located at the frog. Although perceptively heavier, the bow is still playable according to professional violinists. A smaller and slightly lighter prototype is currently under development.



(a). Augmented violin bow



(b). Antenna behind the bridge for position measurements.



(c). The sensing system placed on the bow frog.

Fig. 1. Pictures of the augmented violin prototype

4 Gesture Analysis

We studied three standard types of bow strokes (*Détaché*, *Martelé* and *Spiccato*), by focusing the analysis on accelerometer signals in the bowing direction, which contain the essential information.

4.1 Violin Bow Strokes

Here is a brief description of these bow strokes according to [2].

In *Détaché*, the bow linearly goes from tip to frog and inversely from frog to tip. This linear movement must be adapted to the various dynamics. The bow can be used entirely or in fractions.

Martelé requires a violent gesture. The whole arm must be rapid and vigorous: a very sharp, almost percussive attack must be obtained at each extremity of the bow.

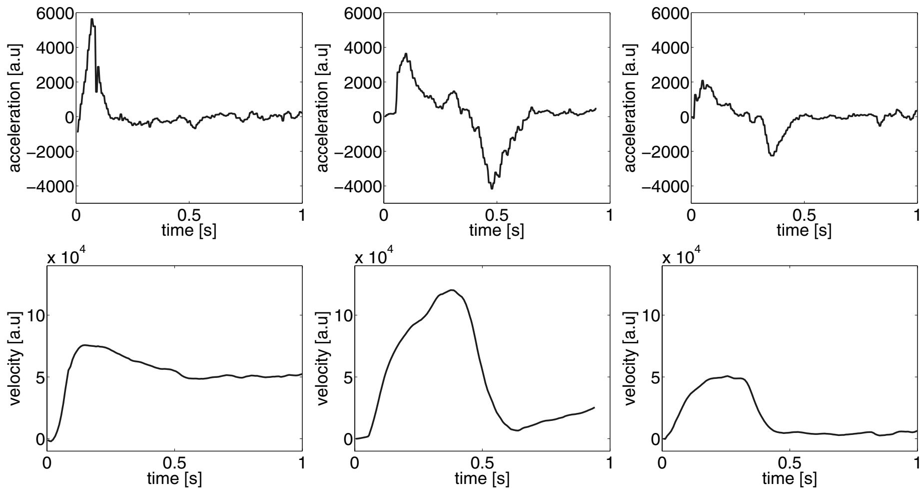
Spiccato uses the phalanges suppleness so that the bow can leave the string after each notes. It results in a light and precise sound.

4.2 Data Acquisition

We built a database from recordings of professional and amateur violinists performing scales in the three bow strokes *Détaché*, *Martelé* and *Spiccato*, at two tempi, 60 bpm and 120 bpm, and three dynamics, *pianissimo* (*pp*), *mezzo forte* (*mf*), *fortissimo* (*ff*).

In order to free the accelerometer signals from angle contributions, we asked the violinists to perform scales on one string at a time and recorded scales on every strings. This way, angle contribution is a constant offset and can be subtracted.

We chose in this study to consider individual strokes. We therefore segmented the recorded gesture data using a peak detection algorithm on the acceleration signals. The gesture database is hence constituted of executions of separate notes,



(a). *Détaché*. (b). *Martelé*. (c). *Spiccato*.

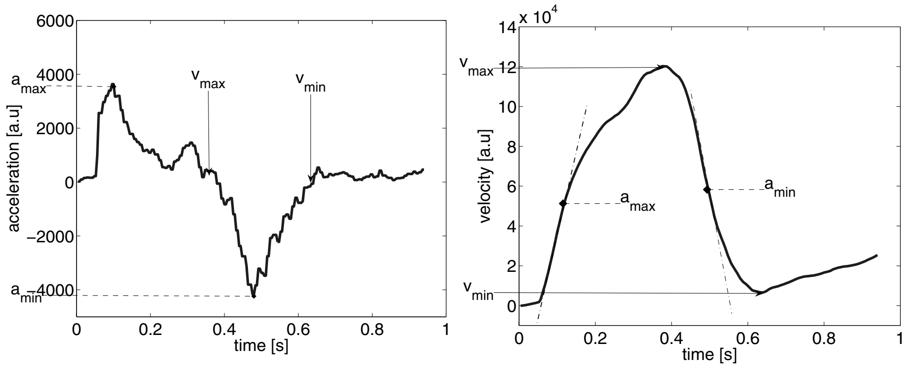
Fig. 2. Acceleration and velocity curves for a single note played in the three styles *Détaché*, *Martelé* and *Spiccato*. Dynamic is *mf* and tempo 60 bpm.

played in three different styles, at three dynamics, two tempi, by two different players.

Figure 2 shows an example of data for the three types of bow strokes *mf* and at 60 bpm. We can see that in *Détaché*, bow velocity remains relatively constant after the attack, unlike *Martelé* and *Spiccato*, where the bow must be slowed down. *Martelé* has typically higher absolute acceleration values compared to *Spiccato*. *Martelé* indeed requires more velocity as it is generally performed using a greater length of bow, compared to *Spiccato*, in order to achieve its typical percussive attack.

4.3 Gesture Features

Four parameters are derived from the acceleration and velocity curves to model the bow strokes: a_{max}, a_{min}, v_{max} and v_{min} (first local minimum after v_{max}), as illustrated on figure 3. Bow velocity is computed from the integration of accelerometers signals. These features correspond to a basic parameterization of the velocity curve shape. They can be computed with sufficient precision and without assuming any model for the velocity shape. They allow for the representation of *Détaché*, *Martelé* and *Spiccato* within a four dimensional space.

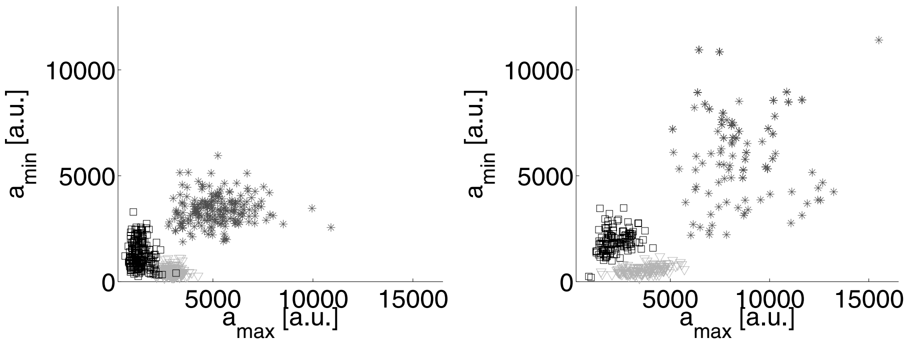


(a). Features on the Acceleration Curve. (b). Features on the Velocity Curve.

Fig. 3. Illustration of the four features a_{max}, a_{min}, v_{max} and v_{min} (first local minimum after v_{max}) on *Martelé* acceleration and velocity curves

4.4 Gesture Space

We used Linear Discriminant Analysis (LDA), which maximizes separation between classes, to estimate the dimensionality of the parameterization. LDA on the gesture database, considering three bow strokes classes, indicates that the class scatter matrix only has two significant eigen values. Therefore, the gesture data can be clustered in a bidimensional space, with maximum in-between classes distance.



(a). Player 1. (b). Player 2.

Fig. 4. Bow Strokes Feature Space (Player Detail). Each point corresponds to a single bow stroke. Fig (a) and (b) show the feature space for each player, at a same dynamic (*mf*) and tempo (60 bpm). Legend is *Détaché* = ∇, *Martelé* = $*$, and *Spiccato* = $\Box$.

We actually found that a_{max} and a_{min}, having major contributions in the eigen vectors, are the two most consistent parameters to model bow strokes, as illustrated in figures 4 and 5. As shown on figures 4(a) and 4(b), for a given dynamic, each bow stroke type forms a separate cluster. Moreover, the disposition of these clusters is similar for both players.

Figure 5(a) illustrates the case where different dynamics are considered. The basic clustering structure remains even if overlap occurs. Nevertheless, for each bow stroke types, sub-structure clustering can be observed as detailed in figures 5(b), 5(c) and 5(d). Precisely, each cluster is composed of three sub-clusters, one for each dynamic variations (*pp*, *mf*, *ff*). *Fortissimo* always corresponds to the highest a_{max} and a_{min} values.

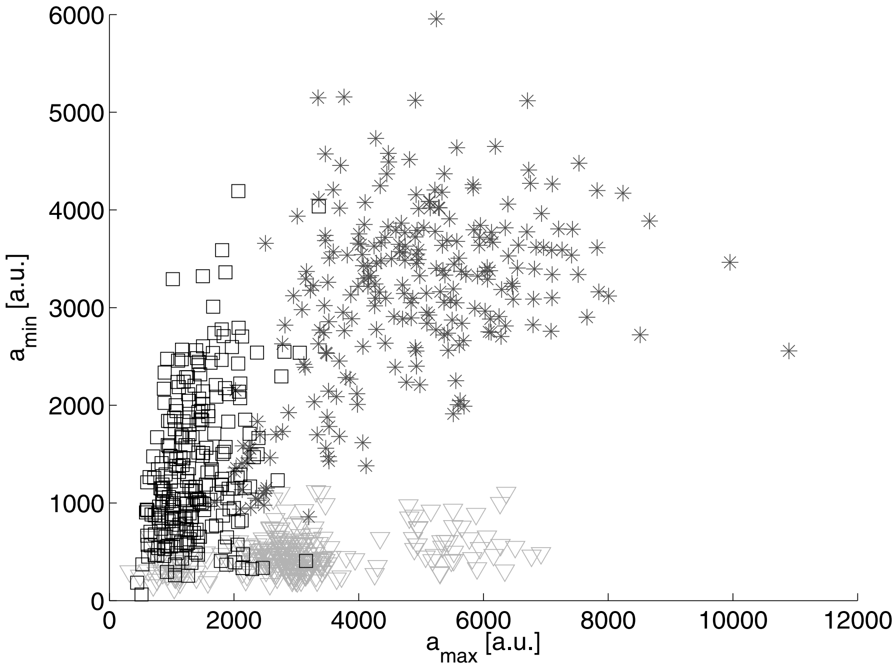
5 Gesture Recognition

We further evaluate the ability of recognizing bow stroke using kNN with a_{max} and a_{min}.Three different test scenarios were chosen.

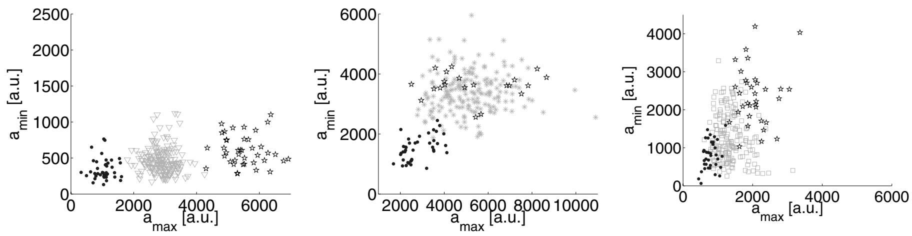
First, we defined three classes, corresponding to the three types of bow strokes. The whole database, i.e. mixing two players, three dynamics and two tempi, is

Table 1. kNN recognition results (Test scenario 1). Database is mixing 2 players, 3 nuances and 2 tempi. Three classes considered.

Test\Ref	*Détaché*	*Martelé*	*Spiccato*
Détaché	**96.7%**	1.3%	2.0%
Martelé	1.0%	**85.8%**	13.2%
Spiccato	6.0%	5.0%	**89.0%**



(a). Bow stroke feature space mixing three dynamics for one player and one tempo (60 bpm). Legend is *Détaché* = ∇, *Martelé* = $*$, and *Spiccato* = $\square$.



(b). *Détaché* cluster: (*pp*= $\bullet$, *mf*= ∇, *ff*= $\star$).

(c). *Martelé* cluster: (*pp*= $\bullet$, *mf*= $*$, *ff*= $\star$).

(d). *Spiccato* cluster: (*pp*= $\bullet$, *mf*= $\square$, *ff*= $\star$).

Fig. 5. Bow Strokes Feature Space (Dynamic Detail). Each point corresponds to a single bow stroke. Fig (a) plots all the features points for one player, at one tempo and at three dynamics. The three bow strokes appear in clusters. Fig (b), (c) and (d) show the detail for each bow stroke cluster: three sub-clusters corresponding to the three dynamics can be seen.

randomly divided into two parts (one-fourth and three-fourths). The quarter of the database, i.e. 320 points, serves as a reference and the remaining three quarter, i.e 1000 points, is used to evaluate the recognition rate. For each test point, vote is done according to the most represented type of bow stroke in the 10 nearest neighbors. Table 1 shows the recognition percentage for this first setup.

In the second scenario, we considered the same three classes but cross-tested the players data: one served as reference for the other. The results are reported in table 2.

Table 2. kNN recognition results (Test scenario 2). Database is mixing 1 player (*Pl1*), 1 nuance (*mf*) and 1 tempo (60bpm). Test points from other player (*Pl2*), same nuance and tempo. Three classes considered.

	Ref	*Pl1*		
Test \		*Det*	*Mar*	*Spi*
Pl2	*Det*	**100.0%**	0.0%	0.0%
	Mar	0.0%	**100.0%**	0.0%
	Spi	6.3%	25.0%	**68.7%**

In the third scenario, we considered each variation of dynamics as a separate class. Thus, nine reference classes, i.e. three types of bow strokes times three nuances for a single player, are tested. This time, two-thirds of the database are used as a reference where each of the nine classes is represented. Table 3 shows the recognition results. For each line, first column is the class of the tested points and the other columns give the percentages of recognition for the nine classes.

Table 3. kNN recognition results (Test scenario 3). Database is mixing 1 player (*Pl1*), 3 nuances, 1 tempo (60bpm). Nine classes considered (3 bow strokes x 3 nuances).

	Ref	*pp*			*mf*			*ff*		
Test \		*Det*	*Mar*	*Spi*	*Det*	*Mar*	*Spi*	*Det*	*Mar*	*Spi*
pp	*Det*	**100.0 %**	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	Mar	0.0 %	**78.6 %**	0.0 %	0.0 %	0.0 %	**7.1 %**	0.0 %	0.0 %	**14.3%**
	Spi	**23.1 %**	0.0 %	**76.9 %**	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
mf	*Det*	0.0 %	**9.5 %**	0.0 %	**90.5 %**	0.0 %	0.0 %	0.0 %	0.0 %	0.0 %
	Mar	0.0 %	0.0 %	0.0 %	0.0 %	**95.8 %**	0.0 %	**4.2 %**	0.0 %	0.0%
	Spi	0.0 %	**35.3 %**	0.0 %	0.0 %	0.0 %	**64.7 %**	0.0 %	0.0 %	0.0 %
ff	*Det*	0.0 %	0.0 %	0.0 %	**6.7 %**	0.0 %	0.0 %	**93.3 %**	0.0 %	0.0 %
	Mar	0.0 %	0.0 %	0.0 %	0.0 %	**85.7 %**	0.0 %	0.0 %	**14.3 %**	0.0%
	Spi	0.0 %	0.0 %	0.0 %	0.0 %	0.0%	**49.9 %**	0.0 %	0.0 %	**50.1%**

6 Discussion

For recognition, three test scenarios were elaborated. The first two scenarios yields high recognition rates. This shows that the three bow strokes are efficiently characterized by the features (a_{min}, a_{max}), even mixing data of two players, different dynamics and tempi, and with a relatively low number of reference data, i.e. one-fourth of the data. Moreover, the cross-player test done in the second scenario confirms the features invariance properties. In this well defined playing situation (scales), our results thus show that the chosen features can be directly related to a music notation level.

In the third scenario, the recognition performances are reduced in some cases. Even with a high proportion of data as reference (two-thirds), confusions occur for example between *Spiccato mf* and *Martelé pp*. However, such confusions are informative as they illustrate actual similarities in bow stroke gestures, when mixing different dynamics. Precisely, from our results, the following different classes, *Spiccato mf*, *Martelé pp* and *Détaché mf*, share similar features, which was actually found to be consistent from the viewpoints of violinists. This shows the limits of recognition approaches since frontiers between classes are not always well defined perceptively.

Furthermore, points that are close in the gesture feature space (figure 5(a)) share similar sound characteristics, e.g. *Martelé pp*, *Détaché mf* and *Spiccato ff*. Consequently, it is perceptually more coherent to characterize bow strokes with a continuous parameterization, using for example a_{max} and a_{min}: such parameters can indeed be related to continuous sound characteristics and/or perceptual features of the listener. It is important to note that a continuous parameterization enables both the recognition of bowing styles and the characterization of hybrid bow strokes.

The results of the study also show that bow acceleration is a parameter of major influence to characterize the different ways of bowing. This comes in complement to acoustic studies on the violin, notably by A. Askenfelt [1] and K. Guettler [5], having already demonstrated the influence of bow acceleration values on the establishments of a Helmhotz regime. It will be interesting to relate the different bowing styles to the number of nominal periods elapsing before Helmholtz triggering occurs, as described in [5].

7 Conclusion and Perspectives

Our goal was to study three different bow strokes, *Détaché*, *Martelé* and *Spiccato*, based on gesture data. After considering basic features based on velocity and acceleration curves, we found that a_{max} and a_{min} provided a pertinent parameterization of these bow strokes. In particular, these parameters enable the recognition of bow stroke types (even in the case of two different players). When considering a higher number of classes including dynamics, we noted typical confusions, consistent with perceptual point of views of violin players and listeners. In summary, our gesture analysis was based on two complementary approaches: recognition and gesture parameterization. Recognition allows us to relate gesture data to music notation, while continuous parameterization of bow strokes could be related to continuous sound characteristics. The detailed relationship between gesture data and sound parameters will be the object of a future study. Moreover, we will investigate other type of parameterizations of the velocity and acceleration that should account for finer characterization of bow strokes. Other parameters such as bow force on strings, pointed by acoustic studies as an influential parameter on sound, will also be considered.

Acknowledgments

We would like to thank Jeanne-Marie Conquer and Hae Sun Kang, violinists from *l'Ensemble Intercontemporain* for their help in this study. We thank Alain Terrier for his contributions in developing the augmented violin prototypes. We also thank René Caussé and Norbert Schnell for support and Suzanne Winsberg for interesting discussions.

References

1. 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.
2. Ami Flammer and Gilles Tordjman. *Le Violon*. J.-C. Lattès and Salabert, 1988.
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. Camille Goudeseune. *Composing with parameters for synthetic instruments*. PhD thesis, University of Illinois Urbana-Champaign, 2001.
5. Knut Guettler. On the creation of the helmholtz motion in bowed strings. *Acta Acustica*, 88, 2002.
6. 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.
7. Joseph A. Paradiso and Neil Gershenfeld. Musical applications of electric field sensing. *Computer Music Journal*, 21(2), 1997.
8. 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.
9. Cornelius Poepel. Synthesized strings for string players. In *Proceedings of the Conference on New Instruments for Musical Expression, NIME*, 2004.
10. 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.
11. Dan Trueman and Perry R. Cook. Bossa: The deconstructed violin reconstructed. In *Proceedings of the International Computer Music Conference, ICMC*, 1999.

12. M.M. Wanderley and M. Battier, editors. *Trends in Gestural Control of Music*. Ircam, 2000.
13. 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.
14. 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.