Combining accelerometer and video camera: Reconstruction of bow velocity profiles

Erwin Schoonderwaldt
CSC
KTH
Lindstedtsv. 24
S-100 44 Stockholm
(+46) 8 790 7561
schoondw@speech.kth.se

Nicolas Rasamimanana
IRCAM
CNRS STMS UMR 9912
Real Time Applications Team
1, Place Igor Stravinsky
75004 Paris
(+33) 1 44 78 48 26
Nicolas.Rasamimanana@ircam.fr

Frédéric Bevilacqua
IRCAM
CNRS STMS UMR 9912
Real Time Applications Team
1, Place Igor Stravinsky
75004 Paris
(+33) 1 44 78 48 31
Frederic.Bevilacqua@ircam.fr

ABSTRACT
A cost-effective method was developed for the estimation of the bow velocity in violin playing, using an accelerometer on the bow in combination with point tracking using a standard video camera. The video data are used to detect the moments of bow direction changes. This information is used for piece-wise integration of the accelerometer signal, resulting in a drift-free reconstructed velocity signal with a high temporal resolution. The method was evaluated using a 3D motion capturing system, providing a reliable reference of the actual bow velocity. The method showed good results when the accelerometer and video stream are synchronized. Additional latency and jitter of the camera stream can importantly decrease the performance of the method, depending on the bow stroke type.

Keywords
Bowing gestures, bowed string, violin, bow velocity, accelerometer, video tracking.

1. INTRODUCTION
Accelerometers and standard video camera are two different types of widely used sensors in the design of cost-effective gesture capture systems. In particular, such sensors have been incorporated in several musical interfaces. Each of these types of sensor has different characteristics. First, accelerometers are typically used to build miniature low latency systems. They are for example particularly well suited to capture percussive gestures. Nevertheless, quantitative use of accelerometer might be difficult due to the fact that the signal depends on both the tilt angle and the actual acceleration. Second, video cameras are well suited to localize, and to spatially follow object. Nevertheless standard video rate are relatively slow for musical application and important latencies are difficult to avoid.

This very brief description points out that accelerometers and video camera are actually complementary systems [8]. Moreover, as discussed by Foxlin [7], there has been a growing interest in the field of Augmented Reality to combine both inertial systems (accelerometers and/or gyroscopes) and vision systems to perform efficient tracking. Generally, the inertial component is fixed to the camera.

We report here a simple approach combining accelerometers with a standard camera to capture bowing gestures. Such a combination is interesting in this case since bowing gesture contains fast and slow temporal features, as well as small and large spatial displacements. As a matter of fact, in bowed string instruments such as the violin, bowing gestures form an essential part in the tone production, giving the player a continuous, yet complex, control of the produced sound.

The capture of bowing gestures have shown important interests in various contexts. Bowing measuring systems can be either used in *non realtime* for fundamental studies of violin playing or as gestural interfaces to control in *realtime* various digital sound processes.

Concerning fundamental studies, the first detailed measurements of bowing parameters were performed by Askenfelt [1, 2], using a bow equipped with sensors for measuring bow force, bow position and bow-bridge distance. Knowledge of how players make use of these parameters provide an important key to violin performance, which could for example be useful for controlling physical models of bowed string instruments or applications in music education.

The *Hyperbow* [6], or more recently the *Augmented violin* [3, 5] are two examples, among others, of modified bows used in live performance. In these particular cases, accelerometers are placed on the bow. The acceleration signal can be used to detect bow stroke and in some cases to recognize bowing styles [5].

Bow velocity is an important parameter to characterize bowing, and is one of the most important input parameters for playing a physical model of the violin. Several acoustic studies ([1, 2]) have also clearly shown the relationship between velocity and sound quality. Nevertheless, bowing velocity can be difficult to measure accurately in a playing situation. Velocity can potentially be derived from video tracking or accelerometer signals. Difficulties arise in both cases:

-Computing accurate velocity profile from video tracking system generally requires the reconstruction of the bowing orientation in 3D space. Moreover, expensive camera for high temporal resolution is generally necessary for fast movements such as bowing attacks. The use of systems such as 3D motion capture system is generally limited to the laboratory environment.

- Computing velocity over a longer time span by integration of the accelerometer signal is problematic. For example, the accelerometer used in the augmented bow [3, 5] is sensitive to both inclination and acceleration (generally referred to as static and dynamic acceleration), which means that there is a variable amount of drift present in the integrated signal.

The method described here, combining the use of both an accelerometer and standard video, allow us to overcome such problems. In particular, we present a method compensating for this drift, enabling therefore the reconstruction of bow velocity from the bow acceleration signal. First, we explain the general principles of the method (section 2). In section 3 the method is assessed using data obtained with a 3D motion capture system, to evaluate the potentiality and limitations of the method. The results are discussed in section 4.

2. BOW VELOCITY RECONSTRUCTION

We first explain the general principle of the setup and reconstruction method. Second, we describe a particular pilot study that serves as a proof-of-principle example.

2.1 Setup

The setup is shown in Fig.1. The two main components are a fixed camera and an accelerometer placed on the bow. The accelerometer axis is set to be parallel to the bow axis. Standard video processing techniques are used to track the bow movements. For example, two color markers can be placed on the bow enabling robust tracking. The accelerometers can be part of the wireless modules described in reference [5] and [6]. Note that simpler implementation is also possible using wired connection between the accelerometer and a sensor A/D interface.

Two angles in this setup are important to consider, both varying significantly during the playing. First, α is the angle between the bow and the vertical direction. The variation of α is the main responsible for the drift in the acceleration signal. Second, β is the angle between the axis of the camera and the direction perpendicular to the bow. The relation between the velocity along the length-axis of the bow and the velocity observed by the camera is then characterized by a scaling factor $cos(\beta)$.

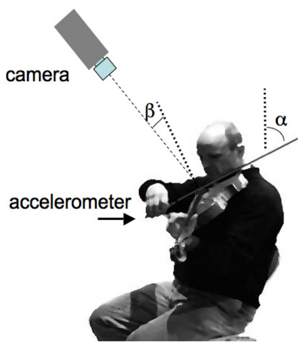

Fig. 1. Setup.

2.2 Computation

As already mentioned, velocity computed from the accelerometer signal typically contains a drift over time. The reconstruction method allows for the compensation of this drift in the integrated signal, by detecting moments when bow direction change. At such moments the bow velocity is equal to zero (referred below as "breakpoints").

Using such information, the accelerometer signal is piece-wise integrated between these "breakpoints". After this step, the velocity curve shows discontinuities at the "breakpoints", which can be removed making an assumption on the form of the acceleration drift. The simplest assumption corresponds to a constant drift in the acceleration signal between two breakpoints, corresponding to a constant α during a bow stroke. This constant drift in the acceleration produces piece-wise constant slopes in the integrated signal. Such linear trends can be simply computed and removed, resulting in a velocity profile that is continuous at the breakpoints.

To obtain the zero-crossings of the bow velocity, a simple video camera can be sufficient. Key points on the bow and the violin are tracked using video processing techniques in order to detect the moments when the bow changes direction. Errors induced by the camera position and the low frame rate of standard video camera are addressed in section 3.

The reconstructed velocity can potentially have a high spatial and temporal resolution (depending on the accelerometer and on A/D conversion system). Therefore, the method offers an easy-to-implement and a cost-effective alternative to expensive commercial motion capture systems to obtain bow velocity signals in violin playing.

The next section describes the implementation of the reconstruction method in a pilot experiment, using a normal video camera and the augmented bow.

2.3 Pilot experiment

A pilot experiment was performed to test the feasibility of the velocity reconstruction method. The bow acceleration was measured with the augmented bow, developed at IRCAM [3, 5]. The bow was equipped with two Analog Device ADXL202 acceleration sensors at the frog, and the acceleration data was sent wirelessly to a RF receiver, connected to a sensor acquisition system, Ethersense [4]. The acceleration data was digitized on 16 bits at the frame rate of 500 Hz.

The video data was obtained with a Sony digital handcam (type DCR-TRV245E). For the measurements reported in Figure 2, two points were marked using differently colored pieces of fabric: one attached to the curl of the violin and the other to the wrist of the player's bowing arm. The camera was positioned in front on the right side of the player, so that the marked points were visible, and the bow motion could be clearly observed (the influence of such setup configuration will be discussed in section 3). The color markers were tracked using Eyesweb software [9], by selecting the pixels with the specified colors and calculating the centre of gravity of the observed pixel regions. The frame rate of the video data was 25 Hz.

The acceleration and video data were synchronized by aligning two synchronization events at the beginning and the end of the recording. The synchronization features were obtained by

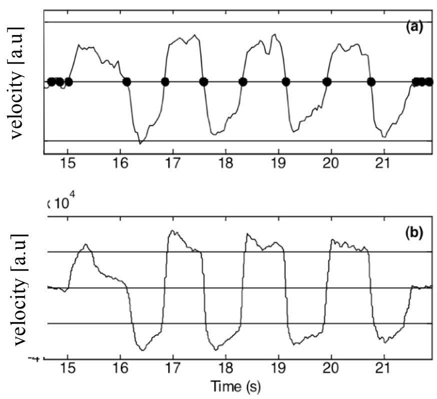

Fig. 2. Détaché bowing on one string. Velocity signal derived from video data (a). The detected zero-crossings are indicated by dots. Reconstructed velocity signal (b) obtained by piece-wise integration of the acceleration of the bow.

tapping with the bow on the curl of the violin, which left a measurable trace in both data series.

From the video data, the position of the wrist marker was obtained in the camera reference. The motion of the violin marker was subtracted to obtain the relative motion of the wrist marker. The derivative was computed and smoothed using Savitzky-Golay filtering (order 5, frame size 11). The zero-crossings of this velocity signal were detected and refined using linear interpolation to obtain a more precise time estimate. Figure 2a shows a series of recorded *détaché* bowing styles: a single tone was repeatedly played to avoid string crossings.

In the next step, the velocity was reconstructed from the acceleration signal of the bow, shown in Figure 2b. The reconstructed velocity profile is coherent as there is an equal repartition between positive and negative velocity, which is in accordance with the performed movement. The reconstructed velocity data has a temporal resolution equal to the accelerometer sampling rate, i.e. 500 Hz. In comparison, the sampling rate of the velocity profile from the video data is only 25 Hz. Moreover, Figure 2 shows that the velocity from the video data is significantly noisier than the velocity profile reconstructed from the accelerometer.

3. METHOD ASSESSMENT

The velocity reconstruction method was further evaluated using 3D motion capture data. During this experiment the bow acceleration was measured simultaneously to the motion of the bow and the violin. The motion capture data provided a reliable reference signal of the actual bow velocity, which was used for quantitative comparison with the reconstructed bow velocity.

The evaluation of the method was performed on three aspects. First, velocity reconstruction is evaluated quantitatively, using an optimal velocity signal. Second, the influence of the video camera position is addressed. Last, the influence of video latency and jitter is quantified.

3.1 Motion capture setup

The motion of the violin and the bow was tracked using a Vicon MX system, with 12 cameras at the frame rate of 150 Hz. The motion capture data was smoothed using Savitzky-Golay filtering (order 3, frame size 9).

Bow acceleration was measured with the same augmented bow described in section 2.3. The sample rate of the acceleration data was 500 Hz. Acceleration data was smoothed using Savitzky-Golay filtering (order 3, frame size 25 – corresponding with the mocap smoothing parameters in the time domain). The data selected for the evaluation was a recording of scales with different bowing styles: *détaché*, *martelé* and *spiccato* played at 60 bpm.

3.2 Velocity reconstruction validation

We first validated the velocity reconstruction method in the case where the velocity zero-crossing were accurately known from the motion capture data, in order to show the achievement of the method under optimal conditions. As the reference signal the velocity along the length-axis of the bow was taken, which corresponds with the actual bow velocity at the string. The reference signal was computed from the motion capture data, compensating for the motion of the violin. During the first validation step the zero-crossings of the reference velocity were used as breakpoints for the reconstruction.

Figure 3 (top) shows the acceleration signal of the bow during a *détaché* scale. It clearly demonstrates the influence of angle α on the acceleration signal, the different offset levels corresponding to playing on different strings.

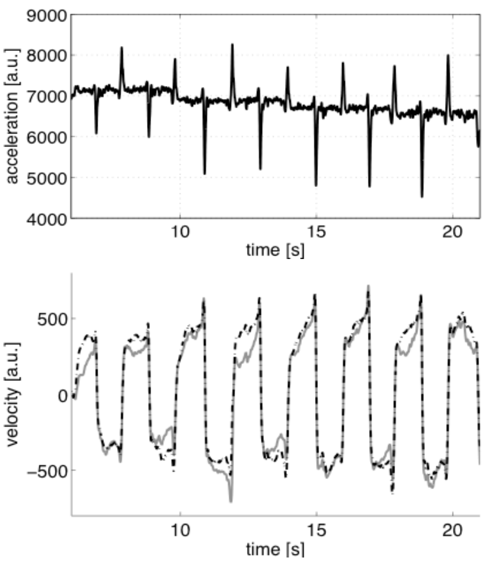

Fig. 3. (top): accelerometer signal during a *détaché* bowing. (bottom): reference velocity (dotted) and reconstructed velocity (plain)

Figure 3 (bottom) shows the reconstructed bow velocity, as well as the bow velocity reference for a series of *détaché* bow strokes using the reference signals zero-crossings as breakpoints. It shows that the reconstructed velocity is in relatively good agreement with the reference velocity. Note that similar results were also found for the other bow stroke types. For example, the correlation coefficients between the reference and the estimated velocities are 0.984 for *détaché*, 0.998 for *martelé* and 0.987 *spiccato* (computed on ascending and descending scale at 60 bpm). These high correlation values demonstrate the validity of the method, when the velocity zeros are accurately determined.

3.3 Influence of the camera viewpoint

The video camera implies a 2D projection of the markers movement. Such a 2D projection is in most cases sufficient since we are not interested in the actual velocity profile from the video data, but only in the determination of the zero-crossing. The most important point is thus to guarantee a sufficiently a high resolution image of the bow movements. Poor resolution can lead to important errors in the determination of the velocity zero-crossing. The optimal viewpoint is therefore the one providing the best overall resolution of the bow displacement.

As mentioned in section 2.1, the projected bow velocity depends on the angle of the camera with the bowing direction with a scaling factor of $\cos(\beta)$. For an optimal projection the video camera should therefore be placed perpendicular to the length-axis of the bow, for example above the player (see Fig. 1).

3.4 Influence of video latency and jitter

Latency and jitter between the accelerometer and video streams could occur, especially if no synchronization mechanism is operated. Such phenomena will mainly affect the timing of detected zero-crossings. In this section, we evaluate the effect of altering such timing in the reconstruction velocity.

An increasing latency was added to the reference breakpoints. The latency varied from 0 to 80 ms by steps of 20 ms. An uniformly distributed random number of span 40 ms was added to simulate a jitter effect. The consequence on the correlation coefficient is shown in Figure 4.

Détaché showed a better robustness to jitter than the two others with a smaller variation. Nevertheless, when the latency is kept sufficiently low, the jitter is not a major source of error for the three tested bowing styles.

For *détaché*, the correlation coefficient decreased from 0.984 to 0.87 when latency increases to 40 ms. However, the correlation coefficient stabilized as the latency increases to 80 ms, dropping only to 0.84. For *martelé*, the method is less sensitive to small latencies. The correlation coefficient is higher than 0.85 for latencies less than 60 ms. S*piccato* appeared to be the more sensitive to latency/jitter effects as the correlation coefficient fell almost linearly.

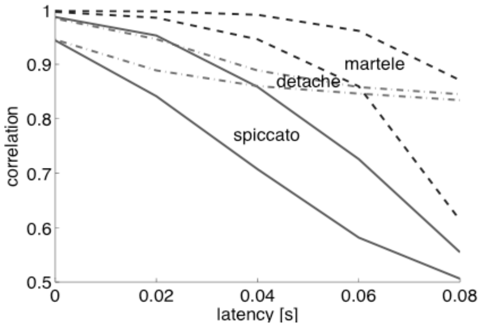

Fig. 4. Correlation factor according to latency and jitter for the three bowing styles. For each bowing style, the upper line is without simulated jitter and lower line with simulated jitter (40 ms).

3.5 Influence of the movement of the player

Except for the errors due to projection and synchronization issues, there are two other possible error sources associated with the reconstruction method. First, as already mentioned it was assumed that the drift between the breakpoints is constant, implying that the angle α in Fig.1 does not change. However, this assumption is not necessary valid in any playing situation, as the player can vary the angle of the bow, especially on the outer G and E strings. Second, the acceleration is measured relative to a fixed reference, rather than the moving violin. This means that movements of the player not directly related to playing are contributing to the reconstructed velocity as well. Thus, the achievement of the method could be dependent on the complexity of the bowing pattern, as well as the amount of additional (expressive) movements by the player.

4. DISCUSSION AND CONCLUSIONS

The overall results show that the method has good potential to reconstruct velocity profile with high temporal resolution. The method is simple to implement and cost-effective compared high-performance motion capture system.

This method could be useful for fundamental studies of bowing gesture, in particular cases where expensive motion capture system systems are not available or too cumbersome to use. The results shown here are promising, but further validation is required to fully characterize the precision and accuracy of this method.

The reconstruction method could also be used in live performance. Nevertheless, the implementation describe here cannot be used as a “strict” realtime system, in the sense that the reconstruction method implies an inherent variable delay. As a matter of fact, the accelerometer drift is not corrected continuously but at discrete time (i.e. velocity zero-crossing). Moreover, as shown in section 3.4, accurate results might require synchronizing the accelerometer and video streams, which would add an additional delay.

Nevertheless, the system can still be useful in performance situation, where detailed information of the bowing gesture is desired and a delayed response is manageable. In such cases, the reconstructed velocity profile can reveal to be very helpful, since the bow velocity is one of the fundamental parameter in the bowing gesture.

The method we described here could also be very valuable for pedagogical applications. For example, accurate information on the playing regularity of specific bow strokes could provide the students with helpful information. In such cases, the information is needed only after the playing.

5. ACKNOWLEDGEMENTS

This work was supported by the Swedish Research Council, project: Bowed String Analysis (621-2001-2537) and the Cost287-ConGas action. We are grateful to Stephane Donikian (IRISA) and Franck Multon (Université Rennes 2) for giving us access to their motion capture system and their precious help during the mocap sessions. We also thank Christophe Desjardins and Emmanuel Fléty for their helpful participations in this project.

6. REFERENCES

[1] Askenfelt, A. *Measurement of bow motion and bow force in violin playing*. J. Acoust. Soc. Am. 80 (4), pp. 1007-1015, 1986.

[2] Askenfelt, A. *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), pp. 503-516, 1989.

[3] Bevilacqua, F., Rasamimanana, N., Fléty, E., Lemouton, S. & Baschet, F. The augmented violin project: case studies, composition and performance report. In *Proceedings of the Conference on New Instruments for Musical Expression*, NIME 2006.

[4] Fléty, E., Leroy, N., Ravarini, J. & Bevilacqua, F. Versatile sensor acquisition system utilizing network technology. In *Proceedings of the Conference on New Instruments for Musical Expression*, NIME, 2004

[5] Rasamimanana, N., Fléty, E., Bevilacqua, F. *Gesture Analysis of Violin Bow Strokes*. GW 2005, Lecture Notes in Artificial Intelligence 3881, S. Gibet, N. Courty, and J.-F. Kamp (Eds.), pp. 145–155, 2006.

[6] Young, D., The hyperbow controller: real-time dynamics measurement of violin performance. In *Proceedings of the 2002 conference on NewInterfaces for Musical Expression*, NIME, Dublin, 2002

[7] Foxlin, E., Naimark, L., "VIS-Tracker: A Wearable Vision-Inertial Self-Tracker". In IEEE Conference on Virtual Reality (VR 2003). Los Angeles, CA.

[8] Kwon, D. Y., Gross, M., Combining Body Sensors and Visual Sensors for Motion Training, Proceedings of ACM SIGCHI ACE 2005, pp. 94-101.

[9] http://www.eyesweb.org