10ème Congrès Français d'Acoustique

Lyon, 12-16 Avril 2010

A haptic simulator for gestural interaction with the bowed string

Stephen Sinclair[1], Jean-Loup Florens[2], Marcelo M. Wanderley[1]

[1] IDMIL, CIRMMT, McGill University, Montreal, sinclair@music.mcgill.ca, marcelo.wanderley@mcgill.ca

[2] ICA-ACROE, INPG, Grenoble, Jean-Loup.Florens@imag.fr

Although human gesture is typically considered to occur at frequencies below 30 Hz, interactions with the environment involve exchange of mechanical energy into the frequency range of sound. These non-linear dynamic situations lend themselves to subtle control over sound-producing phenomena. An example of this is the bowed string. We propose that high-frequency forces experienced during bowing are important factors for accurate control of bowing, providing critical information about the vibration state of the string.

To help test this hypothesis, we have developed a simulator which, using force-feedback hardware, enables synthesis of both sound and friction forces at audio rates. This real-time simulator allows experimentation not only with acoustic parameters, but the presence of haptic feedback also allows examination of how human gestural interaction is affected by model parameters.

We have implemented string models based on two major paradigms: modal synthesis, and the digital waveguide. These are interfaced with the excitation block through a non-linear force-velocity table. These models may be modified to change their fundamental frequency, spectral response, friction characteristics, and body resonance. Using observations of interaction with modified parameters, we plan to develop a sound basis for comparison of different synthesis techniques and their parameters in reference to human gesture.

1 Introduction

Explicit human gesture is a slow phenomenon, generally taking place below 30 Hz [10]. However, when considering interaction with the environment, gestural interaction inevitably involves many forms of non-linear dynamics, which lead to acoustic vibrations occurring at frequencies well into the range of hearing.

A perfect example of where we take advantage of this is in musical interaction. In bowing for instance, the hand moves deliberately only at a slow velocity, but the resulting stick-slip friction produces vibrations well into the audio range. Coupled with a resonator such as a string or bar, it can induce Helmholtz motion to produce harmonic sounds.

In the theory of instrumental interaction [3], it is suggested that while such high-frequency motion goes beyond pure human gesture, the presence of feedback is vital to accurate and precise gestural control over the coupled system. One method of testing this idea has been to use haptic force feedback technology to create synthetic environments which behave according to discrete simulations of physical models, with which a human can interact. Using such a system it is possible to artificially modify the relationship between action and feedback (sound and forces) in ways that would be impossible with a purely mechanical system, in order to study the implications for gestural control.

The system, Fig. 1, is composed of force-feedback hardware, a dedicated sampling and signal processing controller running an algorithm at an adjustable rate, and an extra stereo analog output connected to a loudspeaker or headphones.

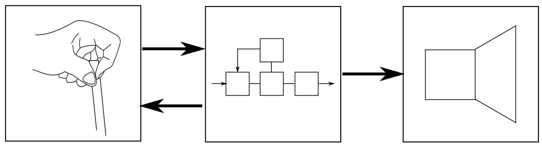

Figure 1: A system diagram showing how the algorithm, force-feedback hardware, and operator are conceptually arranged.

2 Previous work

Acoustic research in bowed string modeling has a long and thorough history. Most of this has focused primarily on the mechanisms of stick-slip friction and string vibration. A branch of this research, mostly in the last two decades, has concentrated on real-time models to allow gesture-based control of sound synthesis. For efficiency, these are typically modeled using either modal synthesis [9, 1] or the 1-D waveguide time-domain synthesis [19].

An advantage of real-time performance vs. off-line simulation is that it allows active exploration and evaluation of the model's parameter space. Rather than limiting to play-back of synthetic or pre-recorded gesture signals, interactive gestures similar to the actual experience of bowing can be used to test the model and locate characteristics of its behaviour [17].

For the bowed string, we feel this is quite important; the non-linear time-variant nature of it suggests that specific combinations of input curves (force, velocity, position, angle) will generate particular behaviours that cannot be reproduced using comparatively static test parameters such as, e.g., the oft-seen "constant velocity" simulation which tend to sound uninteresting compared to real playing. Recently, Demoucron and Maestre have shown that more detailed parametric models of multi-dimensional control gestures based on motion recordings of professional players can indeed improve the quality of sound synthesis of the bowed string [6, 15].

An approach to acquiring gesture signals for control of bowing models is to use an input device that allows integral control in several dimensions. A common choice is the graphic tablet, which provides up to five degrees of freedom from a pen-like interface. This has been used for both digital waveguide [17] and modal [6] synthesis methods.

For a complete simulation, it is necessary to also consider the influence of output forces of the system in addition to the shape of input signals. While human control of *gross* features of input curves can be deliberate, and modeled in a straight-forward manner, forces fed back at the bow-string interaction point generate modulation of the input at higher frequencies.

In general, the behaviour incited by force feedback from the string to the bow has been taken into account in real-time simulations only by simple models of bow dynamics. For example, Adrien suggested the use of a spring-damper system to represent the bow [1]. Such a configuration allows modeling the influence of string-bow coupling on bowing. However, since this feedback does not have a complete channel back to the real-time operator, the model is incomplete from a gestural point of view; energy is effectively absorbed into the spring-damper and then lost, without actually reaching the operator's hand.

It is thus not possible to simulate the full effect of force feedback on gesture using an input-only device, for instance, a graphic tablet. At most, it can simulate relatively smooth gesture curves, but cannot take into account these micro-variations driven by the system resonance, nor can it provide appropriate frictive resistance to movement. Although the result may not noticeably decrease realism of the sound as compared to simulated dynamics, it is not clear how this simplification affects the operator's ability to actually perform real bowing gestures, or to be convinced that he is indeed bowing a string.

Use of a haptic device allows us to enable or disable (or otherwise modify) almost any aspect of the gesture-feedback relationship, to explore these conjectures. However, it is first necessary to determine how to best approximate bowed string interaction using a force feedback device. In this paper we describe two options which have been previously developed, and present an informal comparison.

3 Hardware requirements

A force feedback system allowing to simulate the full range of human-environment interaction must meet some stringent requirements. It is well-known that a high-bandwidth frequency response is required for haptics in order to simulate stiff contact—that is, the device must be able to quickly pass in and out of regions of near-zero impedance to regions of near-infinite impedance [5]. Some additional considerations are maximum force, backdriveability, input and output resolution, and peak acceleration [12].

A fundamental question for our task is whether these performance measures may differ in importance for musical applications. For instance, many such interactions do not involve hard surfaces. Percussion is an exception, in particular with metallic surfaces, but drum membranes have a certain amount of give, as do taught strings. On the other hand, consideration of the frequencies and amplitude of vibration experienced at the interaction point is critical. For vibrations experienced during bowing, the maximum force during "stick" does not exceed about 4 N when bowing close to the frog, averaging closer to 1.5 N normally [16]. However the change in state from stick to slip is instantaneous. Moreover, these state transitions occur at a first harmonic of the played note, therefore in the haptic range of approximately 41 Hz (E1 on a double bass) to outside the haptic range at 660 Hz (E5 on a violin). Evidence suggests that these vibrations are felt by performers, and even for high notes, haptic vibrations are felt during transient portions of a note [4].

We can conclude then that while high maximum force is not necessarily crucial, a haptic system for audio interaction should be: rigid, to avoid distorting the frequency response; able to quickly modulate forces, i.e., have a large dynamic range; and should have high spatial and temporal resolution.

As for inertia, in the general case it is typical to consider that a device should be as transparent as possible, thus low inertia is important. Here, we are attempting to model interaction with a tool (the cello bow), which in reality does have a mass of approximately 50 g. We can therefore get away with a certain amount of inertia in the end effector as long as it can be stimulated at the appropriate frequencies with low distortion.

The hardware used here, called the TGR (*Transducteur Gestuel Rétroactif*) has been designed and built at ACROE, Grenoble, as an experimental apparatus approaching the above requirements. It has been described in previous papers related to the work on modal synthesis described here [8, 14].

The complete ERGOS system, of which TGR is the sensor/actuator component, is described in more detail in [7]. An important component is the controller card, the Toro-16 by Innovative Integration, which uses a Texas Instruments TMS320C6711 digital signal processor to execute audio-haptic algorithms at variable speed, and can handle some algorithms up to the standard audio rate of 44 kHz. This makes it possible to calculate real-time audio and haptic response in a completely synchronous, single-sample manner at high frequencies with 16-bit resolution.

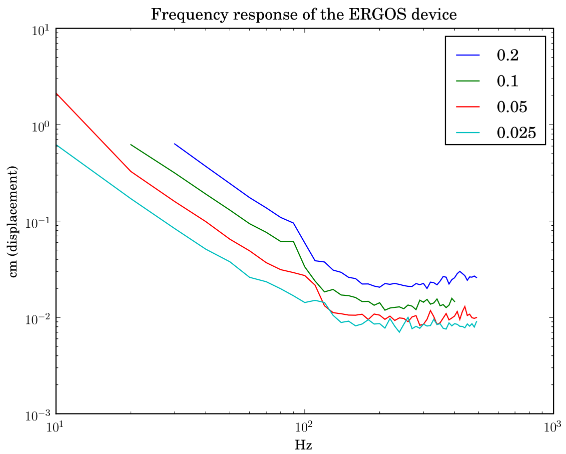

Figure 2: Frequency response of a slice motor of the TGR haptic device. Measurements were taken at intervals of 10 Hz by playing a digitally synthesized sinusoid (at 40 kHz) to the voltage control of the motor at four different amplitudes. The visible cutoff of predictable response at 100 Hz is likely related to the resolution of measurement. The important aspect here is the lack of any obvious resonances. This measurement does not take the 2-DOF coupler attachment into account.

The TGR motors do have a certain mass[1], but are balanced by high force capabilities, giving good acceleration performance. When the coils are active, an analog offset in the force signal is used to compensate gravity. The maximum sustained force of the device is approximately 200 N. Thus if we consider the peak force combined with inertia [12], the acceleration ability of the device is considerable. We have not yet measured it with an accelerometer, but while attempting to measure acceleration with the built-in position sensors we found that it is possible to drive a motor from the center of the workspace to the extremity in less time than one sample when the measurement controller is running at 40 kHz.[2] The driven frequency response of one motor can be seen in Fig. 2. Displacement in this figure was measured using the device's LVDT position sensors sampled with 16-bit resolution.[3]

4 Physical models

Currently two different real-time physical models of the bowed string are implemented on this hardware. The first is based on modal synthesis, while the second uses the 1-D digital waveguide technique.

[1]One motor weighs approximately 150 g, or 330 g for two motors with the 2-DOF attachment.

[2]The workspace of each motor is about 5 cm. The workspace of the 2-DOF end effector attachment is about 13 cm × 5 cm.

[3]It will be necessary in the future to use proper accelerometer-based tools to better characterise the device's frequency response, as differentiating this data introduces significant noise. We also intend to perform clamped isometric measurements to determine force response independent of inertia.

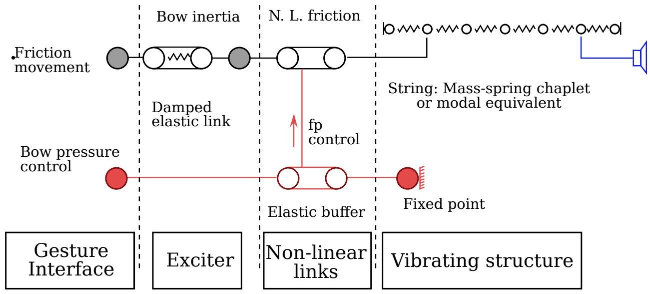

Figure 3: Diagram of the modal synthesis system, from [8].

4.1 Modal synthesis

The modal synthesis is based on a mass-spring model of the string, coupled with a non-linear relationship between velocity and force [8]. The mass-spring model has been transformed to a modal representation with parallel tuned resonators.

Forces resulting from the bow-string interaction are applied to the set of resonating masses which are configured according to the modes of a string. The device end effector is connected to the non-linear link via an elastic buffer in order to provide a stability guarantee. The basic model is thus a mass-spring system operating in two one-dimensional spaces connected through the non-linear link: the resonators and friction model constitute one axis, with another axis providing vertical pressure control through the use of a switched spring ("elastic buffer" in Fig. 3).

4.2 Digital waveguide model

The digital waveguide model is based on the model of the bowed string proposed in [19]. The use of this model for haptic interaction using the ERGOS hardware was described in [18]. A similar concurrent work also proposed a general method for integrating haptic feedback with digital waveguide networks, using the bowed string as a working example [2].

The current system implements haptic interaction via the two-point anchor method described in [18]. This allows coupling the end effector velocity to the string while minimizing noise during low-velocity interaction. Instead of directly taking the velocity of the end effector, the position is followed by an anchor which is only allowed to move according to the stick-slip state of the friction. A second anchor follows more closely to the interaction point, but kept a maximum distance away, similar to the Hayward-Armstrong friction method [11]. The velocity of this second point, which is exactly zero during periods of stiction, is taken to determine transmission of energy between the bow and the string via a non-linear function.

This function, implemented as a look-up table shaped by vertical pressure, is indexed by differential velocity between the bow and the string to determine the string's new velocity. This is inserted into the waveguide, which is split at the interaction junction, and works as the system's resonator. The network and look-up function can be found in Fig. 4.

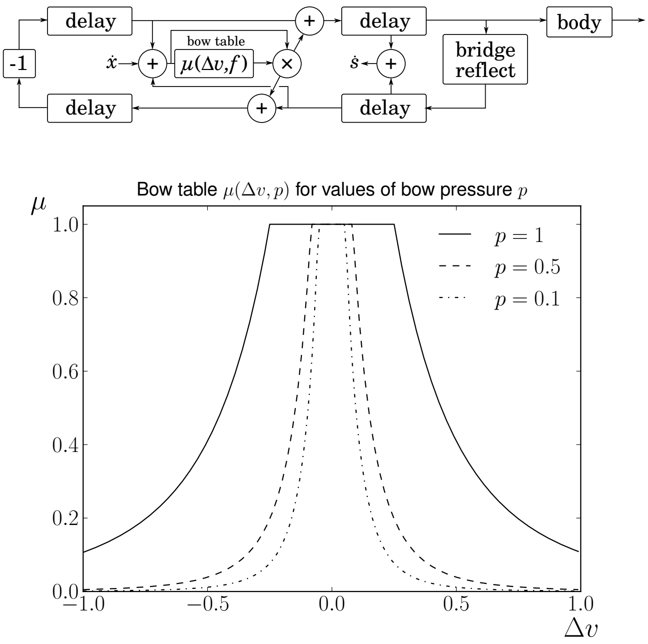

Figure 4: The digital waveguide network used and the non-linear look-up table mapping input differential velocity to new string velocity for several values of bow pressure [19]. The flat portion indicates sticking state.

4.3 Technical limitations

While the Toro-16 card is fairly powerful for signal processing tasks, we have nonetheless found that there are limits to what is possible to execute on it in real time at audio frequencies.

The modal model was designed to execute at 44.1 kHz, and at this rate the maximum number of modes that can be synthesized is 40. For a bass string, this models the signal up to about 500 Hz. This is enough to give a very good impression of the cello timbre, but a real cello contains higher frequencies, particularly when bowed outside the typical playability parameter region. As described shortly, there is clearly more spectral content present in the digital waveguide output.

However, the limitation on the DWG model is sample rate. Using a good velocity estimation technique [13], the maximum sample rate we could establish was 25 kHz, or 24 kHz when including some extra instructions for recording data to a buffer. Using a cheaper velocity estimator, we were able to run the model at 36 kHz, but there was more perceivable noise present.

5 A brief comparison

Although these models are tuned to the same frequency and are executed on the same hardware, a subjective comparison shows some very evident differences. Examples in this section were recorded by writing simulation variables to an on-card memory buffer, transfered after each stroke to the CPU for disk storage. The bow strokes for each model were independent and performed by a human operator. In the DWG recordings, the 500 Hz low-pass "body" filter was removed so that the string velocity signal would be more clearly represented.

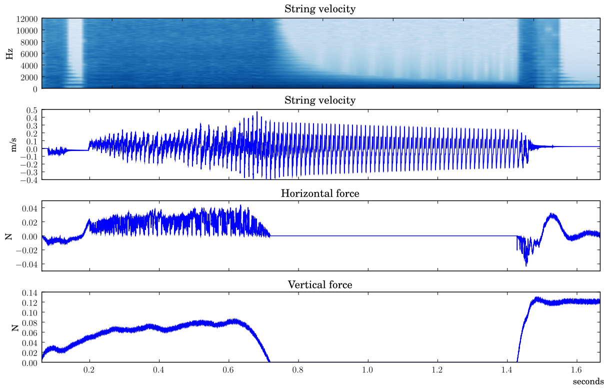

Figure 5: Spectrogram, velocity, and force plots of a bow stroke, actively damped at the end, using the DWG model running at 24 kHz with the body filter removed, string tuned to 80 Hz.

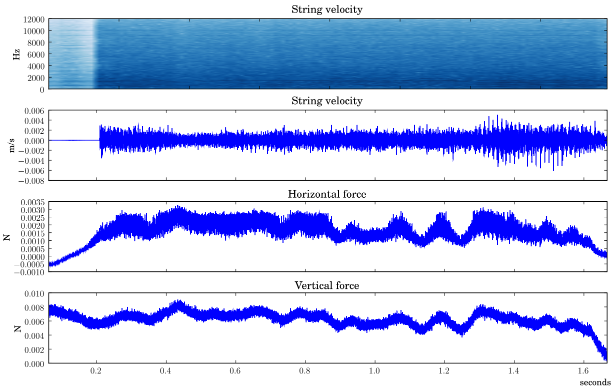

Figure 6: Rasping, appearing as almost pure noise, in the DWG model when touching very lightly.

Foremost, it is clear that more higher modes are synthesized using the waveguide model. (Compare figures 5 and 8.) There is a certain high-frequency response that appears not only through the speaker but also can be heard eminating from the device's motors, an effect which does not occur with the modal model.

When touching very lightly and moving, the waveguide model exhibits a certain "rasping" which is an indication of the presence of noise in the velocity signal (Fig. 6). This noise does not appear when the end effector is at rest, nor do we hear it significantly when vertical pressure is applied. In the modal model, no such noise exists, at least in any obviously perceptible manner, presumably because there are no resonators in the corresponding frequency range. Note that some of this high-frequency content in the DWG model is likely due to aliasing—in particular, no analog anti-aliasing filters are present in the feedback loop.

The sticking response in the waveguide model does not simply apply constant frictive forces opposite the direction of motion as in the modal model, but actually stops the anchor point from moving during stick. As such, when pressing hard the waveguide model exhibits complete stiction, refusing to move entirely. It is possible to correct this unrealistic attribute by limiting the

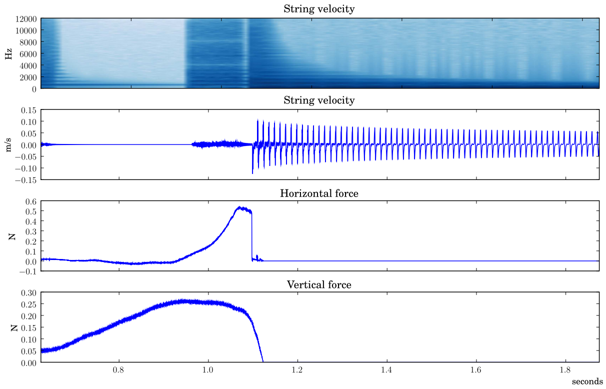

Figure 7: Demonstrating a pluck achieved by pressing hard into the DWG string ($f_0 = 80$ Hz, SR=24 kHz) and pulling horizontally while releasing.

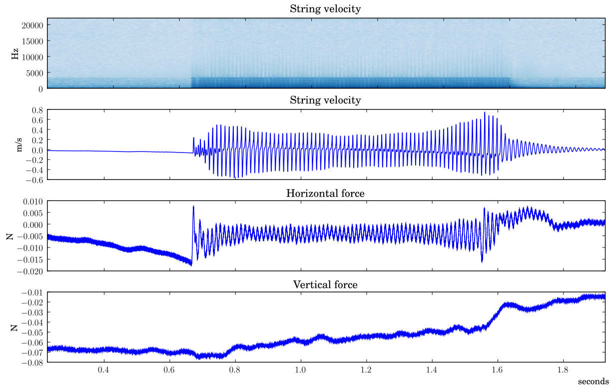

Figure 8: Similar plots for a bow stroke using the modal model, running at 44.1 kHz, string tuned to 80 Hz. Since only 40 modes up to 500 Hz are synthesized, the spectral content is comparatively limited, but also contains much less noise.

effect of vertical pressure, but some testers have found it interesting because it allows a certain "plucking" behaviour to be simulated by pushing and then quickly pulling the end effector away from the string (Fig. 7). Additionally, with moderate force it is possible to get very gritty sounds and feelings, not entirely dissimilar to bowing badly. The modal model does not suffer from this issue because the friction curve is based on forces opposing velocity rather than being based on a spatial derivation of friction as in the H-A friction model.

As an example of how force feedback may affect the system as a whole, we have included a recording of the DWG model with "smooth" friction. Comparing figures 9 and 10, it can be seen that the string oscillates differently in the latter case, while the operator feels very little evidence of it.

6 Conclusion

We have presented a description of hardware and two algorithms for haptic force-feedback interaction with the bowed string and provided an informal comparison of the qualities of both algorithms.

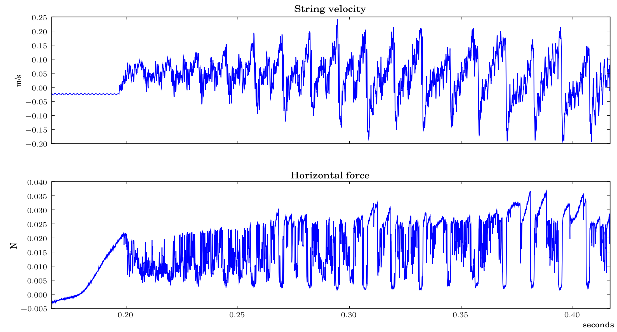

Figure 9: Close-up of the attack for the DWG model running at 24 kHz, with the body filter removed, string tuned to 80 Hz.

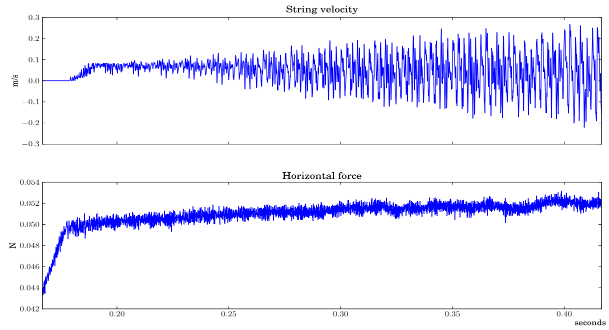

Figure 10: A similar attack in the DWG model, but using a "smooth" friction model—here, horizontal forces experienced by the operator are related only to the direction of movement, not to the string vibration. It can be seen that the string velocity also takes on a different shape.

While we encountered some limitations in terms of execution frequency and number of modes, the overall impression in both cases of bowing a cello string is quite convincing. However, these technical limitations contribute to the perceptual differences between the models. We assume that if the modal model covered the entire frequency spectrum and the DWG model was executing at 44 kHz, the output signals would be much more similar.

In order to perform a future subject-based comparison study on the two models, it will be necessary to either improve on these limitations or reduce one or both models to more closely approximate each other.

For example, a filter could be inserted into the digital waveguide in order to reduce the spectral content. For sound only, this is an obvious possibility, but to reduce spectral content in the force signal one must be careful about introducing filters into the control loop since this can introduce extra delay. Similarly, we might run the modal model at a slower rate, thereby freeing up instruction cycles for more modes.

Another consideration is that 44 kHz (or even 24 kHz) is most likely overkill for the haptic feedback loop. We might increase performance by reducing the haptic sample rate while keeping a high-quality audio rate—this could allow the use of more efficient vector-

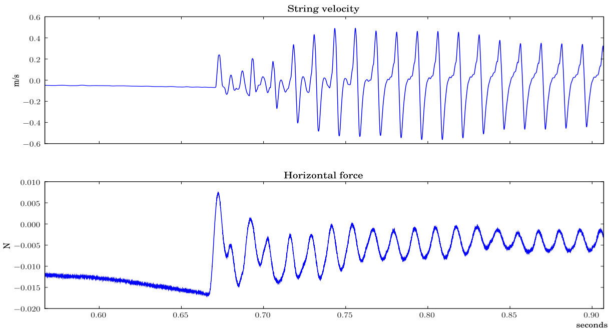

Figure 11: Close-up of an attack in the modal model running at 44.1 kHz, string tuned to 80 Hz.

ized instructions for audio calculations. In fact, in the original work for the modal model [8], the haptic servo was executing at 3000 Hz (synchronously), much lower than the audio rate. The previous paper describing the use of the DWG technique also mentions a multi-rate approach [18].

Lastly, the two bowed string models make use of different (but related) excitation models to create stick-slip motion at the interaction point. However, it is possible to establish methods for interchanging these, for example using the DWG model's two-point anchored spring to drive the modal resonators. This would open more possibilities for comparison, decoupling the excitation and resonator elements of each model.

Funding for this work has been provided in part by the French Agence Nationale de la Recherche project Créativité Instrumentale (ANR-08-CREA-031).

References

[1] J. M. Adrien. *Étude de structures complexes vibrantes, applications à la synthèse par modèle physique.* PhD thesis, Université Paris 6, 1988.

[2] E. Berdahl, G. Niemeyer, and J. Smith, III. Using haptic devices to interface directly with digital waveguide-based musical instruments. In *Proceedings of the Ninth Int. Conf. on New Interfaces for Musical Expression, Pittsburgh, PA*, pages 183–186, 2009.

[3] C. Cadoz. Instrumental gesture and musical composition. In *Proceedings of the 14th Int. Comp. Music Conf.*, pages 1–12, Köln, 1988. ICMA.

[4] C. Chafe. Tactile audio feedback. In *Proceedings of the Int. Comp. Music Conf.*, pages 76–79. ICMA, 1993.

[5] J. E. Colgate and J. M. Brown. Factors affecting the z-width of a haptic display. In *IEEE Conf. on Robotics and Automation*, pages 3205–3210, 1994.

[6] M. Demoucron. *On the control of virtual violins - Physical modelling and control of bowed string instruments.* PhD thesis, Université Pierre et Marie Curie - Paris VI; Royal Institute of Technology, Stockholm, 11 2008.

[7] J. Florens, A. Luciani, C. Cadoz, and N. Castagné. ERGOS: Multi-degrees of freedom and versatile force-feedback panoply. In *Proceedings of Eurohaptics*, pages 356–360, 2004.

[8] J.-L. Florens. Expressive bowing on a virtual string instrument. In A. Camurri and G. Volpe, editors, *Lecture Notes in Comp. Science: Gesture-Based Communication in Human-Comp. Interaction, 5th Int. Gesture Workshop*, volume 2915, pages 487–496. Springer Berlin / Heidelberg, Genova, Italy, April 2003.

[9] J.-L. Florens and C. Cadoz. Modular modelisation and simulation of the instrument. In *Int. Comp. Music Conf.*, Glasgow, 1990.

[10] A. Z. Hajian, D. S. Sanchez, and R. D. Howe. Drum roll: increasing bandwidth through passive impedance modulation. In *Proceedings of the IEEE/RSJ Int. Conf. on Intelligent Robots and Systems. IROS '97*, volume 3, pages 35–36, 1997.

[11] V. Hayward and B. Armstrong. A new computational model of friction applied to haptic rendering. In *Proceedings of The Sixth Int. Symposium on Experimental Robotics VI*, pages 403–412, London, UK, March 2000. Springer Verlag.

[12] V. Hayward and O. R. Astley. Performance measures for haptic interfaces. In G. Giralt and G. Hirzinger, editors, *Robotics Research: The 7th Int. Symposium*, pages 195–207, Berlin, Germany, 1996. Springer Verlag.

[13] F. Janabi-Sharifi, V. Hayward, and C.-S. J. Chen. Discrete-time adaptive windowing for velocity estimation. *IEEE Trans. on Cont. Sys. Tech.*, 8(6):1003–1009, Nov. 2000.

[14] A. Luciani, J.-L. Florens, D. Couroussé, and C. Cadoz. Ergotic sounds: A new way to improve playability, believability and presence of digital musical instruments. In *Proceedings of Enactive '07*, pages 373–376, Grenoble, France, November 2007.

[15] E. Maestre Gómez. *Modeling instrumental gestures: an analysis/synthesis framework for violin bowing.* PhD thesis, Universitat Pompeu Fabra, Spain, 2009.

[16] E. Schoonderwaldt and M. Demoucron. Extraction of bowing parameters from violin performance combining motion capture and sensors. *The Journal of the Acoustical Society of America*, 126(5):2695–2708, 2009.

[17] S. Serafin, R. Dudas, M. M. Wanderley, and X. Rodet. Gestural control of a real-time physical model of a bowed string instrument. In *Proceedings of the Int. Comp. Music Conf.*, pages 375–377, Beijing, China, 1999. ICMA.

[18] S. Sinclair, G. Scavone, and M. M. Wanderley. Audio-haptic interaction with the digital waveguide bowed string. In *Proc. of the Int. Comp. Music Conf.*, pages 275–278, Montreal, Canada, 2009.

[19] J. O. Smith. Efficient simulation of the reed-bore and bow-string mechanisms. In *Proc. of the Int. Comp. Music Conf.*, pages 275–280, The Hague, 1986. CMA.