Perspectives on gesture-sound relationships informed from acoustic instrument studies.

Nicolas Rasamimanana, Florian Kaiser, Frederic Bevilacqua
IRCAM, CNRS - UMR STMS, 1 Place Igor Stravinsky, 75004 PARIS, France
{Nicolas.Rasamimanana, Frederic.Bevilacqua}@ircam.fr, kaiser@nue.tu-berlin.de

Abstract

We present an experimental study on articulation in bowed strings, that provides important elements for a discussion about sound synthesis control. The study focuses on bow acceleration profiles and transient noises, measured for different players for the bowing techniques *detaché* and *martelé*. We found that maximum profiles of these profiles are not synchronous, and temporal shifts are dependant on the bowing techniques. These results allow us to bring out important mechanisms in sound and gesture articulation. In particular, the results reveal potential shortcoming of mapping strategies using simple frame by frame procedures of the data stream. We propose instead to consider input control data as time functions, and consider gesture co-articulation processes.

1. Introduction

When playing music, acoustic musicians face, among others, two types of constraints: physiological and acoustical. These constraints define a range of possibilities that musicians must master to achieve expressive performances.

Similarly to other studies (Leman, 2008), our working hypothesis is that joint investigation of the musician physical movements and the resulting acoustic sound helps to formalize fundamental concepts fruitful for designing digital synthesis control, and more generally approaches in electroacoustic music.

This approach is related to other recent studies on instrumental gesture, in particular studies investigating on the relationships between gesture and sound characteristics in actual playing situations (Goebl, 2004; Dahl, 2000; De Poli et al., 1998;).

In this paper, we focus on string bowing motion and its relationship to sound. Beyond the acoustical point of view, bowing corresponds to a challenging problem considering sound control. First, bowing can be seen as a continuous and simultaneous control of parameters, such as bow speed and pressure to cite the most prominent. Second, bowing can as well produce attacks and articulations, which are also of prime importance. We propose here to particularly pay attention to this second point through the study of bow stroke transitions and their relationships to transient noise. Specifically, we will see that our study points out to limitations of note-based approaches in sound synthesis approach.

Generally, musicians call *articulation* the manner of merging successive notes or more generally groups of notes. On self-sustained musical instruments, like winds or bowed strings, going from one tone to another implies going through a transient phase between two nearly periodic regimes. In the case of bowed strings, this transient phase is characterized by an irregular motion of the string, where one Helmholtz motion stops and a new one develops. Through years of training, string players earn an in-depth control over this transient phase. Adjusting bow main parameters, expert players are able to vary the way transient noise sounds: from smooth and light to harsh and crunchy. From this control, they therefore can produce different kinds of articulations.

Research in music has shown a vivid interest in transient parts of sound, around the idea that they contain key expressive elements. Perception works demonstrated that transients play a predominant role in instrument categorization and recognition (Grey, 1975; McAdams et al., 1995). Sound quality of audio signal synthesis improved thanks to dedicated treatments on transient parts, especially techniques based on signal models (Serra and Smith III, 1990; Dannenberg and Derenyi, 1998; Verma et al., 1997; Roebel, 2003). Acoustic studies investigated the origins of transient parts and their contribution to instrument acoustic signature (Askenfelt, 1993; Goebl et al., 2004). Simulations and experiments on bowing machines were performed and the influence of constant bowing parameters on non-periodic string motions was investigated (Guettler, 2004; Guettler and Askenfelt, 1997; Woodhouse and Galluzzo, 2004). Nevertheless, there are still very few studies on the gestural control of transients.

This paper investigates the relationships between bow movements and sound properties in actual playing situations, paying a particular attention to the temporal behaviours of these multimodal components. The paper is structured as follows. First we recall important concepts on sound and gesture. Second we describe the methodology for specific experiments we performed on bowing articulation. Third, we describe the results. Finally we discuss how these results can provide us with particular perspectives on the control of digital instruments.

2. Sound and gesture articulations in violin

From a sound standpoint, the irregular string motion that occurs during the transient phase results in a broad-band pulsed noise (Chafe, 1990). For bowed string musicians, this typical noise is well-known; they usually learn to control it explicitly or implicitly for expressive purposes. Violin pedagogue Ivan Galamian, talking about sound production on a violin, alludes to this noise saying that percussive sounds like consonants are necessary to shape the melody line formed by the vowel sounds (Galamian, 1999). This sound analogy between bowed string instruments and voice, and more generally between music and speech is also often drawn in music acoustics (Chafe, 1990; Wolfe, 2007; Godoy, 2004). In particular, works led by Wolfe (2002, 2007) investigated clues supporting this analogy and especially brought forward common issues on timing and on energy in voice and instrument sounds. Such comparisons are particularly insightful for the study of players' control on sound articulations.

From the point of view of control movements, bowed string players often consider transitions between strokes as important as strokes themselves. Musicians control continuously bow motion and sound to achieve different expressive cues as shown through the study of bowing techniques in (Rasamimanana et al., 2006) or through the analyses of different performance versions in (De Poli et al., 1998). However, making a transition between two strokes requires as much control skill as sustaining a sound as indicated by violinist Ami Flammer (Flammer and Tordjman, 1988) and can itself be considered as a constitutive part of bowing (Menuhin, 1973).

In bowed string instruments, these two sonic and gestural points of view on articulations are actually summed up in the concept of bowing techniques. Learning and performing bowing techniques indeed concern both sound and gesture. On the one hand, the name of bowing techniques often refer to an "action" (e.g. Martelé “hammered”), on the other hand, the end goal is to achieve sound with specific characteristics (e.g. percussive-like). Qualities of transitions in gesture and sound traditionally result from playing different bowing techniques. However, it may require years for student players to fully master a bowing technique and use it in a musical context. For these reasons, bowing techniques offer a fertile ground and a structured basis for studies on sound and gesture control of string players. In the following, we present a study of two fundamental bowing techniques from a sound and gesture point of view, with the aim to derive more general principles on articulations.

3. Methodology

The methodology followed in this paper is inspired by works of Guettler and Askenfelt (Guettler, 2004; Askenfelt, 1989) on bow transitions. However, our study is aimed towards actual musician performances, instead of controlled acoustics experiments. Our approach is therefore similar to Goebl's study on piano (Goebl et al., 2005). Besides, we analyse the sound as emitted by the whole instrument, instead of the sole string movement. In our case, we take into account the resonance effects from the violin body, closer to the musician perception.

We detail in this section the experimental procedure with a brief description of the considered bowing techniques, the measurement setups and recorded movement parameters, and finally the audio analysis.

3.1 Procedure

Eight violin players participated in the study. They were all advanced level violinists, with 9 to more than 20 years of practice. To measure different articulation qualities, they were asked to perform a one octave, ascending and descending D major scale with the bowing technique *Détaché*, then the same scale with the bowing technique *Martelé*.

Both scales were recorded at a tempo of 80bpm and the dynamic level forte. To reduce possible accelerometer bias due to gravity, subjects were asked to remain on the D string, therefore minimizing the angle variations of the bow. All violinists were asked to perform on the same violin and bow, to guarantee common conditions for all measurements.

3.2 Bowing techniques

Détaché is the most common bowing technique. Each note is performed on a separate bow, hence the name. The sound is kept relatively constant during one stroke and there is no break between notes. In *Détaché,* the articulation corresponds to the transition from one stroke to the other. This transition can be achieved with different degrees of smoothness / harshness but generally remains smoother than *Martelé*. For this kind of stroke, transitions can be compared to liquid consonants such as 'l'. As opposed to *Détaché*, *Martelé* strokes are incisive and sound almost percussive, hence the name. Strokes are generally short, with a harsh beginning and ending. In *Martelé*, the articulation corresponds to the transition between stop (no motion/silence, end of previous stroke) to the beginning of next stroke. Such transitions can be compared to plosive consonants such as 't'. Galamian (1999) actually describes these two bowing techniques as being important poles in bow mastery from which violinists can compose other bowings.

3.3 Bowing measurements

As stated in (Guettler, 2004), bow acceleration is one of the essential parameters influencing sound articulations. Moreover, we previously found in (Rasamimanana et al. ,2006) that bow acceleration is a particularly salient parameter to characterize the two bowing techniques *Détaché* and *Martelé*: differences and similarities between both techniques were characterized using features derived from bow acceleration profiles. Motivated by these previous results, we assume that acceleration is an essential motion parameter for bowing, in particular during bow stroke transitions.

The system used to record players' bowing movements consists in two parts. The first part is a module that measures bow acceleration with two accelerometers (Analog Device ADXL202). This module is mounted at the bow frog with a carbon clip. The placement of the two accelerometers is adjusted to measure bow dynamics along 3 directions: bow stick, strings and orthogonally to the stick. Accelerometer data are digitized at the sampling rate frequency = 333Hz with a resolution of 16 bits and are sent wirelessly with a RF transmitter powered with batteries. This module is shown on Figure 1. The second part consists in a computer interface (Flety et al., 2004) with a dedicated card receiving data from the RF transmitter. Data are sent through ethernet connection to a laptop for recording using the Open Sound Control protocol. Accelerometer data is median filtered with a window of 8 samples to remove eventual acceleration peaks due to HF transmission errors. The total overweight of the system represents 14 grams at the frog: although perceptively heavier, the bow is easily playable according to the subjects. This system is similar to the one used in Bevilacqua et al. (2006).

Figure 1: Module placed at the frog of the violin bow to measure players' bowing movements. It consists of two accelerometers and a RF transmitter, powered with batteries.

3.4 Sound analysis

As explained before, we consider the resulting sound of the whole instrument. To do so, we recorded the violinists' performances with a microphone clipped behind the violin bridge (DPA 4021). Sound is digitized at 44100Hz, in 16 bits. It was recorded simultaneously to acceleration data using Max / MSP.

The sound analysis consists in extracting transient noise. The general approach is to use signal processing techniques assuming a signal model with deterministic and stochastic components. The extraction of the transient noise can then be performed using analysis / synthesis techniques. The general procedure is to estimate the parameters of a signal model describing the deterministic components (analysis) and generate a new signal on the basis of this model (synthesis).

Subtracting this modelled signal from the original signal, we get a residual signal that contains the transient noise. We subsequently estimate the quantity of transient noise by computing the energy of the residual.

Because of the short time span of transient parts, between 50 ms and 90 ms in playing situations (Guettler, 1997), the chosen model in this paper is based on the formalism of High Resolution Methods (HRM). In HRM, the deterministic components are modelled as exponentially modulated sinusoids. This actually gives HRM a higher frequency resolution than Fourier especially on short windows therefore enabling a more precise estimation of sinusoid parameters. The method applied in this paper is based on previous works on the use of High Resolution Methods in audio signal analysis (Badeau, 2005; Laroche, 1989), using ESPRIT for the estimation of the sinusoid parameters (Badeau et al., 2005), (see Annex for details).

4. Experimental results on bowing

Recorded waveforms and bow acceleration are shown on Figure 2 for a series of strokes performed as *Détaché* (top) and *Martelé* (bottom). The residual energy profile, corresponding to transient noise, is also plotted (see computation details in Annex). As expected, in case of *Détaché,* the transient noise peaks lie at transitions between strokes. In the case of *Martelé*, the transient noise peaks are mainly located at the start and end of strokes, which correspond to moments when periodic string vibrations are initiated and stopped. Moreover, as already noted in a previous study on similar bowstrokes (Rasamimanana et al., 2006), each *Détaché* stroke is characterized by one acceleration peak, while two acceleration peaks (acceleration and deceleration) occur in each *Martelé* stroke.

For statistical analysis, we built a dataset by isolating bow articulations for each of the bowing techniques. The segmentation is performed in two steps. First, we achieve a manual segmentation to select instants corresponding to articulations in the sound files. Second, an automatic process adjusts the segments limits based on the transient noise and the acceleration profiles: limits are determined by the energy and acceleration standard deviations. Vertical dotted lines delimiting the analysis segments are shown in Figure 2.

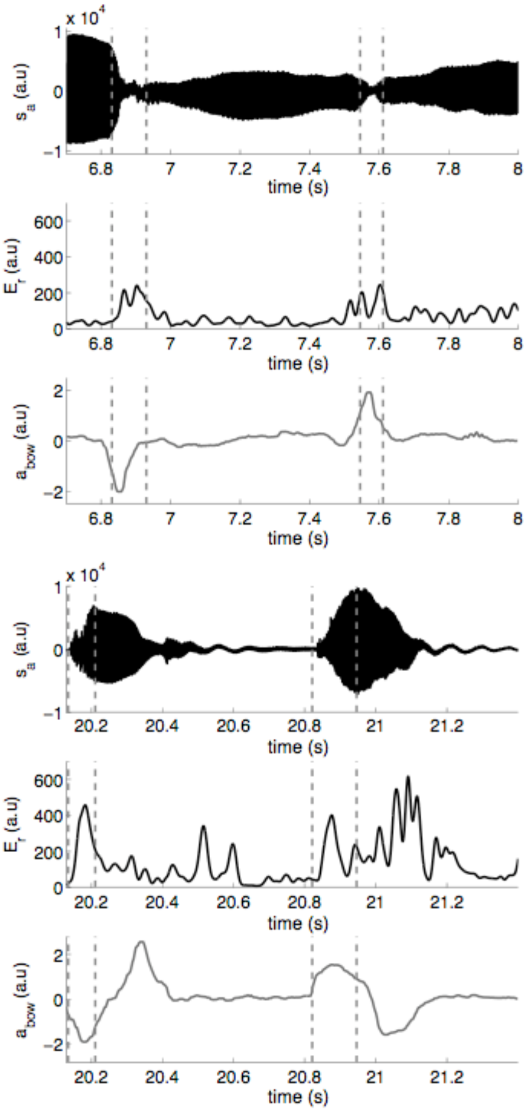

Figure 2: From top to bottom: *Détaché* audio signal waveform, residual energy (transient noise), bow acceleration, and *Martelé* audio signal waveform, residual energy (transient noise), bow acceleration. Vertical bars delimit the analysis segments.

For both gesture and sound, we observe that each articulation presents specific temporal distributions, as shown in Figure 3 for *Détaché* and *Martelé*. Note that for clarity, the distributions are normalized (maximum is equalled to one). For each articulation, acceleration and transient noise profiles exhibit different bell shapes. Interestingly we can observe small time shifts between the two profiles, which actually varies on the bowing techniques. To quantitatively assess these shifts, the first order moment of the profiles are computed for acceleration and transient noise.

$$t_m = \frac{\sum s_n * t_n}{\sum s_n},$$

where s_n and t_n are respectively the digitized signal and the samples index.

The time shift Δtm, quantified by the difference between moments of the residual energy distribution and acceleration distribution ($\Delta t_m = t_m^{residual} - t_m^{acceleration}$), is found to be positive for *Détaché* and negative for *Martelé*. We further examined such temporal features in different cases and players.

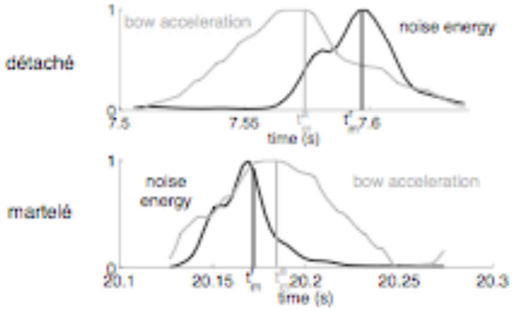

Figure 3: Normalized temporal distributions of residual energy (dark), representing the sound transient noise, and acceleration absolute value (light) for one articulation in Détaché (top) and one articulation in Martelé (bottom). trm and tam respectively designate first order moments for residual energy and bow acceleration.

Figure 4 shows the succession of articulations on the scale exercise for both bowing techniques *Détaché* and *Martelé* (player 8). All thirteen stroke transitions in the scale played *Détaché* show positive time shifts, while the fourteen strokes in *Martelé* show negative time shifts. Precisely, the ensemble of *Détaché* transitions is characterized by a Δtm median value of 15ms and an interquartile of 12.3ms, while for *Martelé* the Δtm median value is -18ms with an interquartile of 10.8ms. These values show that there is a statistically relevant timing difference between the two bowing techniques. These values also seem to reveal that the different bowing techniques imply distinct motion-sound relationships.

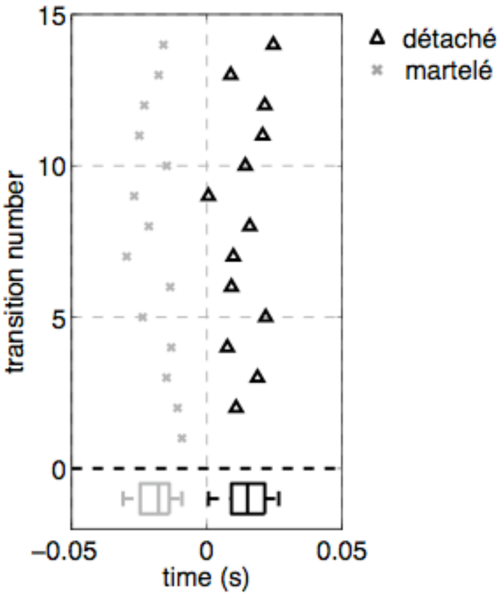

Figure 4: Δtm computed for one violinist articulations on the scales in Détaché (dark Δ) and Martelé (light X). Each symbol corresponds to a stroke transitions. Boxplots give synthetic views for each scale.

We now extend the analysis to eight violin players. In spite of player idiosyncrasies, we can see that the average time shifts Δtm remains positive for *Détaché* and negative for *Martelé* , as shown on Figure 5. Quantitatively, over all players, *Détaché* articulations are characterized by a Δtm median value of 19ms and an interquantile of 15ms. *Martelé* articulations are characterized by a Δtm median value of -20ms and an interquantile of 21ms. This actually confirms on a broader statistical level that temporal motion-sound relationships can be specifically related to articulation types. As expected, some inter-player variability can be found. Such variability could be interpreted as possible differences in articulation "pronunciations": some players "uttered" tones in a globally more distinct way.

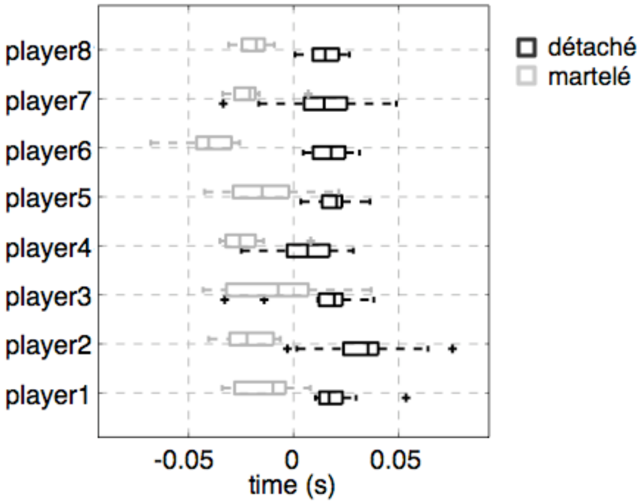

Figure 5: Δtm for eight violin players for each scale in Détaché (dark) and Martelé (light). Each boxplot represents one acending - descending scale.

5. Discussion and perspectives on sound control

We presented an experimental study on violin articulations, focusing on both bowing motion and sound properties. Particularly, bow acceleration and noise component of the sound were measured. These parameters constitute key elements for the sound control in violin playing. We especially looked into note and gesture transitions, which have been understudied. Interestingly, under this scope, we found that bow acceleration and "noise" are not in a direct causal relationship, even if acceleration is recognized as an important acoustic parameter that directly influences transient noise (Guettler, 2004; Woodhouse and Galuzzo, 2004). Precisely, we found that transient noise could appear either before or after acceleration peaks and this actual time offset can consistently depend on particular bowing techniques. This can be partially understood considering that transient noise for *détaché* and *martelé* always peaks after the note onset. However, in upbow-dowbown *detaché,* the acceleration peaks appear exactly between two separate continuous strokes, while it appears slightly after the attack of *martelé.*

Thus, the role of past gesture is fundamental for the correct interpretation of the acceleration data. This aspect can be regarded as gesture co-articulation (Rasamimanana, 2008, Rasamimanana and Bevilacqua, to appear).

Of course, a complete physical model including all parameters, (e.g. at least complete bow position, velocity and pressure temporal profiles) could explain these results. Nevertheless, our point here is *time* relationships between control and sound parameters are complex. Considering possible consequence on mapping strategies, our results show that simple strategies directly linking motion values to sound parameters on a frame-to-frame basis could not replicate the type of articulations considered in this study. To avoid such shortcoming, it seems important to consider the approach, schematically illustrated in Figure 6a, which is a generalisation of our experimental Figure 3. This figure illustrates that we need to separate, at the signal level, a raw gesture level and the actual signal used at the synthesis level. Importantly, our point is here to propose explicitly a temporal approach in the transformation between these two levels. Each phrase is transformed through a specific temporal process (such as time convolution in case of linear process). These temporal processes overlap and depend on both the previous, current and forthcoming process. This principle is illustrated by comparing the Figures 6a and 6b, where the order of the gesture sequence is changed: A, B, C in Figure 6a and C, B, A in Figure 6b. The effect of this permutation should fundamentally change the morphology of the sound objects. We elaborate on this approach with the three points below that seem to us essential in sound synthesis control to overcome the limitations of a note approach.

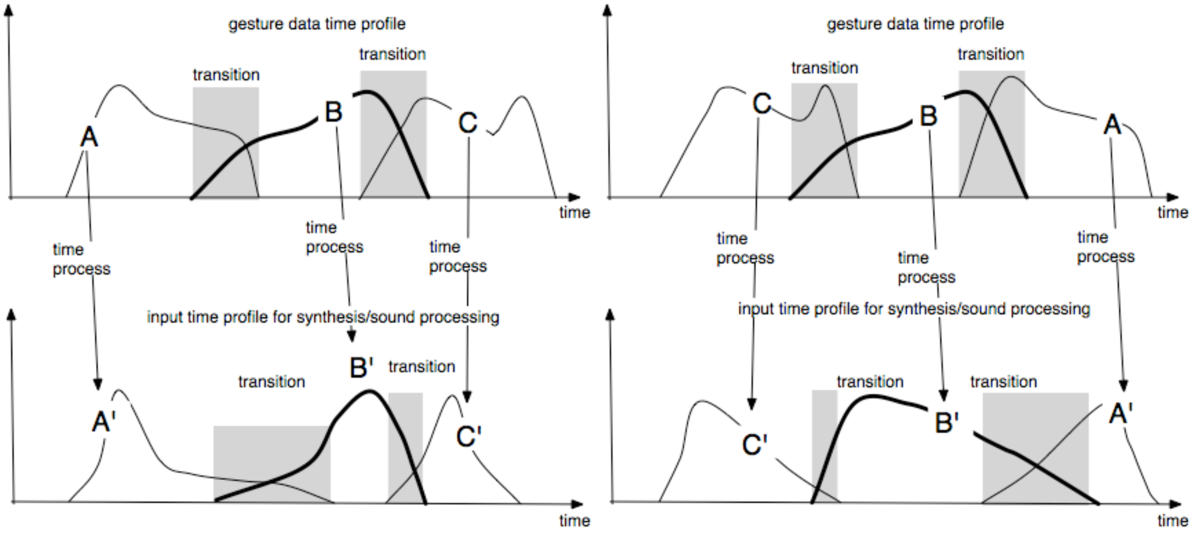

Figure 6: Temporal mapping schema: gesture data time profiles are transformed to input control profiles for sound synthesis. Gesture sequence order is reversed between (a) and (b): the morphology of the sound objects is changed as an effect of the gesture permutation.

First, gesture data should be considered as temporal functions instead of a stream of data. As a matter of fact, gesture data are most often processed as individual data frame, as this is directly handled in several programming environment (e.g. Max, Pd).

For example, the MIDI protocol is clearly based on a “note approach” with a single parameter to model the attack (velocity). If continuous control can be achieved through the use of “aftertouch” parameters, complex articulations cannot be properly reproduced. Second, the control parameters should consider both the previous and following notes.
Such a phenomena is analogue to co-articulation found in speech, and we propose here that its transposition to gesture, known as gesture co-articulation (Ortmann, 1929; Palmer, 2006; Rasamimanana and Bevilacqua, to appear), should be considered. Transitions cannot be simply taken into account note based approaches such as MIDI for example. Third, gesture to sound mapping should contain intrinsic dynamic time behaviour. Such mechanisms, as intrinsically incorporated in physical modelling (Henry, 2004), could model adequately the types of temporal shifts found in our results. Such a model can actually encompass fine articulation mechanisms as found in bowing.
Our recent work on gesture following (Bevilacqua, 2007) could incorporate these different aspects. In this processing system, the time profiles of gesture data are analysed. Time index of these input profiles can be then put in correspondence to other profiles that are the actual input for sound synthesis, as illustrated in Figure 6. These two levels of time profiles can be set either manually or using appropriate algorithms. Compared to other mapping strategies that operate principally on spatial relationships, the mapping strategy we propose is therefore in the time domain, and could take into account articulations as measured in the study reported here. Such “temporal” mappings are currently being experimented with.

6. Acknowledgements

The authors would like to acknowledge Roland Badeau for his help on High Resolution Methods. The authors also thank Matthias Demoucron, Julien Bloit, Norbert Schnell, and René Caussé for fruitful discussions and support. The authors particularly thank Anne Mercier, Florence Baschet and the students of the Dijon Music Conservatory for their participation in this study. This work has been partially supported by the European Commission 7th Framework Programme SAME project (no. 215749)

ANNEX

A. High Resolution model and estimation

The model mostly used derives from Fourier spectral analysis where the deterministic components are represented as a sum of sinusoids with variable amplitudes, frequencies and phases. Because of the short time span of transient parts, between 50ms and 90ms in playing situations (Guettler and Askenfelt, 1997), the chosen model in this paper is based on the formalism of High Resolution Methods (HRM). In HRM, the deterministic components are modelled as exponentially modulated sinusoids. This actually give HRM a higher frequency resolution than Fourier especially on short windows therefore enabling a more precise estimation of sinusoid parameters. The method applied in this paper is based on previous works on the use of High Resolution

Methods in audio signal analysis (Badeau, 2005; Laroche, 1989), using ESPRIT for the estimation of the sinusoid parameters (Badeau et al., 2005).

A.1 Signal model

The deterministic components are modelled as a sum of exponentially modulated sinusoids.

For all t $\in \mathbb{Z}$,

$$s(t) = \sum_{j=1}^{K} \alpha_k z_k^t \tag{1}$$

where K $\in \mathbb{N}^*$ is the order of the model. $\alpha_k \in \mathbb{C}^*$ are complex amplitudes. $z_k \in \mathbb{C}^*$ are distinct complex poles. z_k can be written as $z_k = e^{\delta_k} e^{2i\pi f_k}$, with $\delta_k \in \mathbb{R}$ the sinusoid modulating factor and $f_k \in \mathbb{R}$ the sinusoid frequency.

The observed signal is then represented as the combination of the deterministic component model s(t) and an independent, centered, white gaussian noise w(t) with a given variance sigma.

$$x(t) = s(t) + w(t)$$

A.2 Parameter estimation

In this paper, the estimation of the parameters, i.e. amplitudes and poles, is based on a property of the modelled signal covariance matrix Rss(t): the rank of Rss(t) is exactly K, the number of distinct poles, if it is of size n > K and computed from l > K observations. This has a direct consequence on the observed signal covariance matrix [Eq 5] : a study on its rank permits to separate the observed signal space into two orthogonal subspaces, the signal space spanned by the exponentially modulated sinusoids and its orthogonal complementary, the noise space. Namely, the eigenvalues of the observed signal covariance matrix Rxx(t) are

$$R_{xx}(t) = R_{ss}(t) + \sigma^2 I_n$$

The poles are computed using the K first eigenvectors of Rxx(t) combined with the property that the signal space is actually spanned by the poles. This is done with the ESPRIT algorithm (Badeau et al., 2005), based on the rotational invariance property of the signal space. The poles amplitudes are finally estimated with a least square regression.

A.3 Application to a violin recording

Previous studies showed that the ESPRIT algorithm provides an accurate estimation of the frequency of the deterministic components under the condition of an additive white noise (Badeau, 2005). To optimize the performance of the parameter estimation, the recorded audio signals are cut into eight frequency subbands of equal width. The analysis is then carried out independently on each subband, assuming a constant noise power on each of them. The window size used to perform the analysis is 128 samples at Fsa = 44100Hz, i.e. 2ms. The number of exponentially modulated sinusoids K is usually unknown, although it plays a key role in the algorithm performances. For this study, K is set to 20 sinusoids per subbands. This value actually overestimates the theoretical value of 18 for D4 (294Hz), but ensures a correct estimation of the poles and their amplitudes (Laroche, 1989).

References

Askenfelt, A. (1989). Measurement of the bowing parameters in violin playing. ii: bow-bridge distance, dynamic range, and limits of bow force. *The Journal of the Acoustical Society of America*, 86(2).

Askenfelt, A. (1993). Observations on the transient components of the piano tone. *In Proceedings of the Stockholm Music Acoustics Conference (SMAC)*, volume 79, 297-301.

Badeau, R. (2005). *Méthodes à haute résolution pour l'estimation et le suivi de sinusoides modulées*. Application aux signaux de musique. PhD thesis, Ecole Nationale Supérieure des Télécommunications.

Badeau, R., Richard, G., and David, B. (2005). Fast adaptive esprit algorithm. *In IEEE Workshop on Statistical Signal Processing SSP'05*, Bordeaux, France.

Bevilacqua, F., Rasamimanana, N. H., Fléty, E., Lemouton, S., and Baschet, F. (2006). The augmented violin project: research, composition and performance report. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.

Bevilacqua, F., Guedy, F., Fléty, E., Leroy, N. and Schnell, N. (2007). Wireless sensor interface and gesture-follower for music pedagogy*. In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.

Chafe, C. (1990). Pulsed noise in self-sustained oscillations of musical instruments. *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, volume 2, 1157-1160.

S. Dahl. The playing of an accent - preliminary observations from temporal and kinematic analysis of percussionists. *Journal of New Music Research*, 29:225–233, 2000

Dannenberg, R. and Derenyi, I. (1998). Combining instrument and performance models for high-quality music synthesis. *Journal of New Music Research,* 27(3):211-238.

De Poli, G., Rodà, A., and Vidolin, A. (1998). Note-by-note analysis of the influence of expressive intentions and musical structure in violin performance. *The Journal of New Music Research*, 27(3):293-321.

Demoucron, M., Askenfelt, A., and Caussé, R. (2008). Observations on bow changes in violin performance. *In Proceedings of Acoustics*.

Flammer, A. and Tordjman, G. (1988). *Le Violon*. J.-C. Lattès and Salabert.

Fléty, E., Leroy, N., Ravarini, J.-C., and Bevilacqua, F. (2004). Versatile sensor acquisition system utilizing network technology. *In Proceedings of the International Conference on New Interfaces for Musical Expression (NIME)*.

Galamian, I. (1999). *Principles of Violin Playing and Teaching*. Shar Products Co.

Godoy, R. I. (2004). Gestural imagery in the service of musical imagery. In Camurri, A. and Volpe, G., editors, *Lecture Notes in Artificial Intelligence, LNAI 2915*, 55-62. Springer Verlag.

Goebl, W., Bresin, R., and Galembo, A. (2004). Once again: The perception of piano touch and tone. can touch audibly change piano sound independently of intensity? *In Proceedings of the International Symposium on Musical Acoustics*.

Goebl, W., Bresin, R., and Galembo, A. (2005). Touch and temporal behavior of grand piano actions. *The Journal of the Acoustical Society of America*, 118(2):1154-1165.

Grey, J. M. (1975). *An exploration of musical timbre using computer-based techniques for analysis, synthesis and perceptual scaling*. PhD thesis, Stanford University.

Guettler, K. (2004). Looking at starting transients and tone coloring of the bowed string. *In Proceedings of Frontiers of Research on Speech and Music*.

Guettler, K. and Askenfelt, A. (1997). Acceptance limits for the duration of pre-helmholtz transients in bowed string attacks. *The Journal of the Acoustical Society of America,* 101(5):2903-2913.

Henry, C. (2004). Physical Modeling for Pure Data (pmpd) and real time interation with an audio synthesis. *In Proceedings of the Sound and Music Computing Conference*, SMC.

Laroche, J. (1989). A new analysis/synthesis system of musical signals using prony's method-application to heavily damped percussive sounds. *In Proceedings of the International Conference on Acoustics, Speech, and Signal Processing, ICASSP*, volume 3, 2053-2056.

Leman, M. (2008). *Embodied music congnition and mediation technology*. MIT Press.

McAdams, S., Winsberg, S., de Soete, G., and Krimphoff, J. (1995). Perceptual scaling of synthesized musical timbres: common dimensions, specficities and latent subject classes. *Psychological Research*, 58:177-192.

Menuhin, Y. (1973). *L'Art de jouer du violon*. Buchet / Chastel.

O. Ortmann. *The physiological mechanics of piano technique*. Dutton, New York, 1929.

C. Palmer. Nature of memory for music performance skills. In *Music, Motor Control and the Brain*. Oxford University Press, 2006.

Rasamimanana, N. H., Flety, E., and Bevilacqua, F. (2006). Gesture analysis of violin bow strokes. *In Gesture in Human-Computer Interaction and Simulation, Lecture Notes in Computer Science / Artificial Intelligence (LNAI)*, volume 3881, 145-155. Springer Verlag.

Rasamimanana, N. H. (2008). *Geste instrumental du violoniste en situation de jeu: analyse et modélisation. Violin player instrumental gesture: analysis and modelling*. Université Paris 6, IRCAM UMR STMS.

Rasamimanana, N. H., and Bevilacqua, F. (to appear). Effort-based analysisof bowing movements: evidence of anticipation effects. *Journal of New Music Research*.

Röbel, A. (2003). A new approach to transient processing in the phase vocoder, *In proceedings of the 6th International Conference on Digital Audio Effects (DAFx'03).*

Serra, X. and Smith III, J. O. (1990). Spectral modeling synthesis: A sound analysis/synthesis system based on a deterministic plus stochastic decomposition. *Computer Music Journal*, 14(4):12-24.

Verma, T., Levine, S., and Meng, T. (1997). Transient modeling synthesis: a flexible analysis/synthesis tool for transient signals. *In Proceedings of the International Computer Music Conference*.

Wolfe, J. (2002). Speech and music, acoustics and coding, and what music might be 'for'. *In Proceedings of the 7th International Conference on Music Perception and Cognition, Sydney*.

Wolfe, J. (2007). Speech and music: acoustics, signals and the relation between them. *In Proceedings of the International Conference on Music Communication Science*.

Woodhouse, J. and Galluzzo, P. M. (2004). The bowed string as we know it today. *Acustica - Acta Acustica*, 90:579 - 589.