REAL-TIME BOWED STRING FEATURE EXTRACTION FOR PERFORMANCE APPLICATIONS

Kurijn Buys
Queen Mary University of London
k.buys@qmul.ac.uk

Andrew McPherson
Queen Mary University of London
a.mcpherson@qmul.ac.uk

ABSTRACT

Although the physics of the bowed violin string are well understood, most audio feature extraction algorithms for violin still rely on general-purpose signal processing methods with latencies and accuracy rates that are unsuitable for real-time professional-calibre performance. Starting from a pickup which cleanly captures the motion of the bowed string with minimal colouration from the bridge and body, we present a lightweight time-domain method for modelling string motion using segmented linear regression. The algorithm leverages knowledge of the patterns of Helmholtz motion to produce a set of features which can be used for control of real-time synthesis processes. The goal of the paper is not a back-extraction of physical ground truth, but a responsive, low-latency feature space suitable for performance applications.

1. INTRODUCTION

The current paper discusses an implementation of a real-time feature extraction algorithm on the violin, designed specifically for time-domain bowed string signals. The aim is to develop digital instruments which use acoustic instrument signals as control information for digital synthesis. Besides the preservation of ergonomics and physiological stimuli, this approach enables the new sounds to be closely related to the response of the acoustic instrument, so that the player's learned musical skills still apply.

This approach has been explored and evaluated in various studies in the last two decades, with Jehan's Audio-Driven Timbre Synthesizer [1], Essl and O'Modhrain identifying augmented instruments and "reappropriation of instrumental gestures" as ways to repurpose existing musical expertise [2], Janer and Maestre's voice-driven sound synthesis [3], Tremblay and Schwarz's concept of "recycling virtuosity" [4], Puckette's "grafting" of synthesis patches onto live musical instruments [5], McMillen's StringPort [6] and Poepel's PhD thesis on audio signal-driven sound synthesis [7].

Another approach to repurpose player skills is to use a musical instrument controller with an ergonomically similar behaviour as an original instrument, which is e.g. studied in [8]. However, even though physical models could be used to regenerate similar sounds from the gesture data provided by such a controller, it is known that the vibrational behaviour of musical instruments in relation to the playing gestures is very complex, so that the instrument would feel unfamiliar to the player. Present-day physical models of the violin for instance, are still not capable of reliably predicting the transient details to a given bow gesture [9]. This is in particular so for the control over the wide variety of subtle sound differences, to which professional players have invested many years of training.

In addition to Poepel, McMillen and others, we not only seek to inform new sound generation using generic audio descriptors, but to rely on the underlying physics of the bowed string and the related perceptual features, to obtain specific instrument-related features. We hypothesise that introducing this instrument-awareness in the design of the algorithm enables a more efficient and precise implementation than using generic sound analysis, and more suitable to reduce the full audio data to preserve principal components that are meaningful to the player's musical expression. Meanwhile, a much richer and less discrete feature set is obtained than only MIDI data for instance, so that the subtleties of the expression are captured.

That being said, our goal is not to have a physically-accurate set of features (such as the the bowing-parameters, which can be indirectly acquired using statistical audio analysis [10]) but one with subjectively relevant qualities, so we purposely use an empirical approximation, as opposed to exact physical theories. The interest of this approach is also underlined in paradigms such as Cook's "Physically Informed Sonic Modelling" [11], or Farnell's "Procedural Audio" [12]. It is shown in this paper how a piecewise-linear signal model enables the extraction of extremely precise sound features, capturing a substantial part of the sound properties.

The work presented in this paper forms part of a project with a broader scope, aimed at the design of new musical instruments that repurpose the expertise of professional musicians who are known to play with a traditional instrument. As with many other human motor tasks, it is known that a musical instrument player's focus is on the target, that is, the sound, while their gestures are produced by an internal (sensorimotor) mental model that couples action to perception [13]. In other words, the player is not actively thinking about their gestures in a musical performance and the instrument becomes, so to speak, transparent to the performer [14]. Hence, to aid the task of repurposing expertise it is a reasonable idea to promote a familiarity in the

musical interaction by using the sound of a traditional instrument to inform the generation of new sounds.

2. MOTION OF THE BOWED STRING

It is a reasonable assumption that the bowed string vibration captures almost all the player's musical intentions, even though the violin body resonances are disregarded [7, p. 164]. This is so, as the body and sound radiation can be interpreted as filters on an escaping fraction of the string energy. As a result, any intended sound output must be noticeable in the original oscillatory mechanism, i.e. in the string vibration. In this paper we focus on the periodic regime. Non-periodic regimes (occurring at some note onsets or e.g. when the bow force is too high) will be considered in later work.

2.1 Signal-model Of Oscillation Cycles

In the late nineteenth century, Helmholtz discovered that the bowed string vibrates in a V-shape, with the corner traveling up and down the string terminations [15]. Later Raman developed a theory that could explain the ideal Helmholtz motion, assuming a lossless string model with reflection coefficients less than unity and a bow-string friction coefficient relative to the sliding velocity between the string and the bow [16].

In the time domain, this lossless and non-stiff string displacement is a triangular signal, where the corner position relative to its neighbouring corners corresponds to the relative measurement position along the string [17]. However, in practice the interplay of dissipation and stiffness in the string and the characteristics of the bow-string interaction can direct this position somewhat towards the middle (this can be understood for instance in terms of a reduction of higher harmonics in the Fourier series of a sawtooth wave, which results in a smoothening of the discontinuity in that signal, i.e. a relocation of the corner towards the middle). These losses also cause the otherwise perfectly sharp corners to become rounded, reducing the high frequency content from the sound, a phenomenon that has been first described by Cremer, who proposed an empirical model with smoothed corners [18]. McIntyre further extended this model to include transient behaviour, which lead to the "digital waveguide model" of bowed-string motion [19]. In contrast to the "steady-state" regime where the oscillations are perfectly repetitive cycles, the oscillations in the transient regimes are non-repetitive, which is the result of an energy imbalance between the excitation (injecting energy into the resonator) and the resonator (collecting and consuming that energy in a standing wave); in its turn caused by a change of the physical input parameters such as the bow speed or force, or a finger pressing the string. The energy storage property of the resonator and the nonlinear character of the bow-string interaction will seek for a new consensus for that given input parameter state, which demands a short settling time. Given that there are relatively few losses in a string resonator, and the steep stick-slip transition of the bow-string interaction relatively quickly introduces new energy in the resonator,

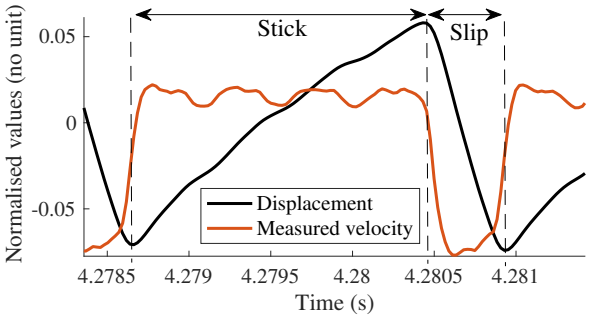

Figure 1. Example of a measured velocity cycle and the corresponding displacement signal with indication of the stick and slip regions.

the attack transients are fast while decays are slow. Many other effects can be distinguished in the vibrational signal of the string, such as "(multiple) flyback", "multiple slip", "Schelleng ripples", "bow-string scraping noise", etc. For an overview see [9].

2.2 Measurement

One significant advantage of using the string vibration signal is that it represents a remarkably simple geometric shape (due to the phase-locking property of the bow-string interaction), which enables a relevant time-domain use of the measured string vibration for real-time analysis, so that cumbersome and intrusive frequency domain methods can be avoided.

To minimise the effects of bridge and body resonances, it is ideal to measure the string signal on the string itself. It was concluded that the "StringAmp" pick-up system [1] was a suitable system for this task. This system uses the strings as electric conductors moving in a magnetic field created by magnets placed under the fingerboard. We developed a custom preamplifier that allows quadraphonic capture of the violin strings. Hence, the pick-up system produces a signal proportional to the velocity of the string at its measured location, near the fingerboard end at the side of the bridge. The displacement signal can be easily obtained from the velocity via numerical integration. Furthermore, a high-pass filter is used to prevent the amplification of DC-offset and low frequency noise from the bowing gestures. Appropriate cut-off frequencies are found to be in the range of 15 Hz to 50 Hz. Figure 1 shows an example of a measured velocity oscillation cycle along with the corresponding displacement signal.

3. SEGMENTATION ALGORITHM

The theoretical findings presented in section 2.1 and our own empirical observations of string displacement data lead us to the idea to design of an algorithm that first approximates the measured displacement signal by an ideal Helmholtz signal. While this model only captures an amplitude, fundamental frequency and relative corner position,

[1] produced by MusikLab Danemark, http://www.stringamp.com

these features can be obtained directly from each oscillation cycle with sub-sample precision while their time-variation provides transient effects. We hypothesise that this information relevantly captures important aspects of the player's musical intent.

Given that the Helmholtz motion in the periodic oscillation regime predicts a cyclic linearly up-and down going displacement signal, it was concluded that a linear segmented signal model, consisting of two line segments per cycle, would be an appropriate first model. The mathematical challenge with this model is to produce a regression of line segments that minimises the total residual error. In a first stage (developed in section 3.1), our algorithm identifies these lines-segmented cycles, resulting in initial break-points, i.e. estimated time values of the Helmholtz corners. In a second stage (detailed in section 3.2) the break-points between the segments are optimised with an iterative regression. Finally, in a third stage (explained in section 3.3) features are extracted from the regressed model and from its relation to the original data.

3.1 Initial Segmentation Estimation

Given that our algorithm is considering Helmholtz motion of the string, a fast and computationally lightweight time-domain detection of the period is possible. In this first stage, crude estimations of the Helmholtz corners are identified. The algorithm consists of detecting signal changes from a positive to a negative RMS threshold and vice versa. By using thresholds greater than zero, false identifications by small additional oscillations such as "flybacks" or "Schelleng ripples" (see [9]) are avoided. Finally, the location of the minimum and maximum values between the up and down transitions are used as initial "break-point" values (i.e. the time values of the transitions between the linear segments).

It was empirically found that the displacement signal is optimal for this task, since its oscillation cycles are most clearly distinguishable. It should be noted however that this method requires the RMS calculations to accurately follow the amplitude of the oscillation, which is so, provided that the low-pass filter used in the RMS calculation is appropriately chosen so that it is slow enough to be independent of the signal oscillations, but fast enough to follow transient behaviour. Therefore, on one hand, its cut-off frequency should be sufficiently below the fundamental frequency of the lowest note of interest, while on the other hand, it should ideally be higher than the frequency of the amplitude changes caused by transient behaviour. In practice, a frequency of 150 Hz is found to be suitable for the detection of all four strings. Only in rare particular cases does this lead to erroneous results. However, this is not the focus of this paper and alternative initial segmentation strategies are envisaged in later work.

Figure 2 shows the positive and negative RMS curves (in dashed blue) calculated for two measured oscillation cycles, as well as the initially segmented estimation (in dashed green).

3.2 Optimisation Through Regression

This algorithm optimises the initially provided break-points by minimising the mean square error between the linear segmented model and the data, where the break-points represent non-linear parameters. The purpose of this procedure is to optimally fit the entire cycle in two line segments, which represents an approximation that is maximally reliable with regard to all data provided to avoid ambiguous approximations and to improve precision. Since the effect of any moved break-point influences the regression of neighbouring segments and therefore could influence any break-point value in theory, the ideal regression would require the entire signal to be known in advance. However, this would not allow for a real-time implementation. Hence, a compromise is made where a number of N segments is isolated, i.e. assuming the limiting break-points Φ_k and Φ_{k+N} to be fixed (where k is a arbitrary first-break-point index). When that regression is completed, a new set of N segments is chosen by removing the oldest segment (on the left) and including a new segment (on the right), i.e. with limiting break-points Φ_{k+1} and Φ_{k+N+1}. While the rightmost regressions will be influenced by the imprecision of the initial break-point that is held fixed, the leftmost segments become independent of that effect.

3.2.1 Segmented Regression model

The chosen approach draws on a method proposed by Muggeo, which enables a probabilistic linear segmented fit on a set of discrete data using an iterated optimisation of initially chosen break-points [20]. Muggeo's regression model is based on a discrete set of observations, say $\{Z_n, Y_n\}$ [2] with no knowledge about the relation between those observations (as is generally the case for regression problems). However, in the current case of discretely sampled audio data with a relevant sampling frequency, it is known that intermediate data can be approximated by interpolation. The inclusion of interpolation enables an increased fractional sample precision and avoids conversion to false local minima (since samples at break-point values can partially belong to each of the two surrounding segments). It can be shown that a regression including an increasing amount of interpolated data converges to using continuous interpolated functions of the explanatory variables $Z = Z_n + z'$ and response variables $Y = Y_n + y'_n$ (where $z'([0, 1[)$ and $y'_n([0, 1[)$ represent the continuous interpolation functions over one sample), and replacing the summation over the observations by an integral.

To enforce a regression with a fixed leftmost break-point, say Φ_k, the regression must be expressed as relative from that break-point and its ordinate α_k (updated after each regression), which can be achieved by subtraction of it. The newly defined explanatory variable-function is $z = Z - \Phi_k$, the break-points are $\varphi_i = \Phi_{k+i} - \Phi_k$ (note that $\varphi_0 = 0$) and the response function becomes $y(z) = Y(Z) - \alpha_k$, but for notational convenience the z argument will be dropped after the introduction of a variable, or where it should be clear from the context. The data concerned

[2] note that an index n always refer to a sample number while indexes $i, j, ...$ or numeric indexes refer to a break-point number.

in this regression is comprised in the restricted domain $z = [0, \varphi_N[$, where φ_N is the rightmost (fixed) break-point.

The continuous time model equation can be simply obtained by replacing the discrete with the continuous variables:

$$y(z) = \beta_0 z_0 + \sum_{i=1}^{N-1} \beta_i (z - \varphi_i)_+ + r(z) \quad (1)$$

where β_0 is the slope of the first segment and the succeeding β_i are the differential segment slopes (i.e. how much change in slope there is compared to the previous segment), i.e. $\sum_{i=0}^{m} \beta_i$ is the slope of segment m, with $i = \{1, 2, ..., N-1\}$ and for $i < j$, $\varphi_i < \varphi_j < \varphi_N$ and r is a residual signal. Hence, the segmented regression model y_m, to be used in this iterative procedure at each step s, is

$$y = y_m^{(s)} + r^{(s)} = \beta_0 z_0 + \sum_{i=1}^{N-1} \left(\beta_i z_i^{(s)} + \gamma_i c_i^{(s)} \right) + r^{(s)} \quad (2)$$

where $\{\beta_0, \beta_i, \gamma_i\}$ are the model's parameters and

$$z_i^{(s)} = \left(z - \varphi_i^{(s)} \right)_+ = \begin{cases} 0 & z \leq \varphi_i^{(s)} \\ z - \varphi_i^{(s)} & \varphi_i^{(s)} < z \end{cases} \quad (3)$$

$$c_i^{(s)} = (-1)\, I\left(z > \varphi_i^{(s)} \right) = \begin{cases} 0 & z \leq \varphi_i^{(s)} \\ -1 & \varphi_i^{(s)} < z \end{cases} \quad (4)$$

are the explanatory variables. After regression, the break-point values can be updated as follows

$$\varphi_i^{(s+1)} = \varphi_i^{(s)} + \frac{\gamma_i}{\beta_i} \quad (5)$$

And iterated regressions are performed until desirable conversion. Note that hereafter the iteration step argument (s) will be left out for notational convenience.

3.2.2 Minimising mean square error

To find an optimal fit, first the sum of the least square error can be expressed

$$R^2 = \int_0^{\varphi_N} (y - y_m)^2 \, \mathrm{d}z \quad (6)$$

and its minima with regard to all parameters are found by looking for the zeros of the partial derivatives with respect to the parameters. E.g. for β_j this gives

$$\frac{\partial R^2}{\partial \beta_j} = \int_0^{\varphi_N} \Bigg[2\beta_j z_j^2 - 2 y z_j + 2\beta_0 z_0 z_j + 2 \sum_{\substack{i=1 \\ i \neq j}}^{N-1} (\beta_i z_i z_j + \gamma_i z_j c_i) \Bigg] \mathrm{d}z = 0 \quad (7)$$

This results is a set of $2N - 1$ equations and unknowns, which can be solved after the integration is calculated for each term. For terms of the form $z_i c_{j \geq i}$, $z_{j \geq i} c_i$ and $c_i c_j$, not containing any response variable, the linear interpolation results in linear functions, which allows to analytically reduce the calculations to only include the lower and upper limits of the integration. For instance the terms $\int_0^{\varphi_N} z_i z_{j>i} \, \mathrm{d}z$ reduce to

$$\frac{\varphi_N^3 - \varphi_j^3}{3} - (\varphi_j + \varphi_i) \frac{\varphi_N^2 - \varphi_j^2}{2} + \varphi_i \varphi_j (\varphi_N - \varphi_j) \quad (8)$$

and analogous expressions can be found for the other terms. Terms containing a response variable require the interpolation equation to be introduced. Writing $y = y_n + a_n z'$ with sample-slopes $a_n = y_{n+1} - y_n$, the integral over z becomes a sum of integrals over each interpolated sample $\sum_{n=n_i}^{n_N} \int_{f_{i,n}}^{f_{N,n}} \mathrm{d}z'$ with $n_i = \lfloor \varphi_i - z_{n=0} \rfloor$ the sample index of the sample where break-point φ_i occurs, and where

$$f_{i,n} = \begin{cases} 1 & n < n_i \\ \varphi_i - z_n & n = n_i \\ 0 & n_i < n \end{cases} \quad (9)$$

is the fractional part of each sample to be included in the integration (for each φ_i). Hence, the $\sum_{n=n_i}^{n_N} \int_{f_{i,n}}^{f_{N,n}} y z_i \, \mathrm{d}z'$ terms reduce to

$$\sum_{n=n_i}^{n_N} \Bigg[a_n \frac{f_{N,n}^3 - f_{i,n}^3}{3} + (a_n z_{i,n} + y_n) \frac{f_{N,n}^2 - f_{i,n}^2}{2} + y_n (z_n - \varphi_i)(f_{N,n} - f_{i,n}) \Bigg] \quad (10)$$

and an analogous expression can be found for terms of the form $y c_i$.

Once the β_j and γ_j coefficients are obtained, it can be verified if γ_j is inferior to a tolerance value γ_{th} of choice, which ends the iteration if true for all j and otherwise triggers a new regression with updated break-points using equation (5). Figure 2 shows an example of a few oscillations with an initial segmentation (in dashed green) and regressed segmented estimations (in solid red).

3.2.3 Exception Handling

After each regression, apart from the convergence condition $\gamma_j < \gamma_{\text{th}}\ \forall j$, two more conditions are verified to handle singularities and other exceptions.

While most initially detected segments converge reasonably quickly, there are some exceptional cases. Box stipulates that convergence is not necessarily guaranteed with this linearised iterative regression model [21]. Therefore, the initial break-points should be chosen in a convex region around the finally regressed break-point, while the iteration updates should not lead the break-points out of that region. As already reported by Muggeo, in certain cases the algorithm can stagnate in alternating between two values, so that the tolerance value γ_{th} is never reached [22]. In other cases, an alternation with eventual conversion occurs, but often only after a long iterative process. For now, we simply allow a maximum number of iterations that is

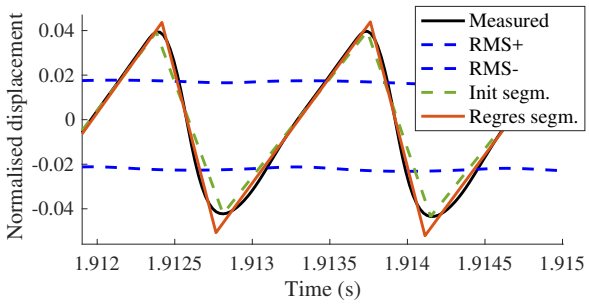

Figure 2. Example of a few measured steady-state oscillation cycles (in solid black) produced with an up bow on the E-string, playing an F# (744 Hz). The blue dashed lines represent the positive and negative RMS signals, the dashed green line shows the initial segmentation and the solid red line shows the regressed segmentation.

sufficient for almost all tested data. It was noted that exceptions exceeding 100 iterations only occur for particular wave-shapes, which is not further discussed in this paper.

The foremost exception occurs when the a set of successive break-points lose the time-order when recalculated after an iteration. This can occur when the data between the break-points is significantly conflicting with the model, i.e. when it is significantly nonlinear in a way that it misleads the first order Taylor series approximation, leading to an excessive break-point jump. For such cases, a reasonable workaround was found by shifting the new break-point values closer and closer towards the values from the previous iteration (repeatedly calculating the mean value) until the time-order conflict is resolved. This exception typically occurs when the data and initially estimated break-points do not conform with the model, e.g, for some multiple-slip, and transient regimes. Since we plan to improve the initial segmentation strategies for this purpose, we are not concerned about these exceptions at this stage.

3.3 Feature Extraction

As mentioned earlier, the acoustically-informed segmentation model is expected to provide characteristics that are inherent to the sound which in its turn captures the player's musical intentions.

The features can be extracted from the parameters of a single complete oscillation cycle, which is defined as one stick and one slip, or two line segments encompassing three breakpoints.

3.3.1 Amplitude and fundamental frequency

Provided that the oscillation cycles were identified correctly in the initial segmentation stage, the amplitude and the fundamental frequency can be derived directly from the current model by respectively calculating the signal amplitude between the successive break-points and the temporal distance between each pair of segments. Multiple consecutive periods could be considered, equivalent to a moving average filter, to reduce noise in the parameters at the cost of latency.

3.3.2 Relative corner position

The model also enables retrieval of a third cycle-related feature, given by the position of the middle break-point within each oscillation cycle (i.e. the corner position relative to the oscillation cycle, e.g. in figure 1 the ratio of the stick duration over the stick and slip durations together), which will be referred to as the "(relative) corner position" or the "duty cycle" [3] .

This feature bears an interesting relation to physically known aspects. While the acoustic losses and string stiffness try to bring the duty cycle to 50% (as explained in subsection 2.1), it is the bow-string-interaction that tends to force the duty cycle towards the ideal Helmholtz case [17]. Hence, it may be expected that a bow release from the string could be identified by a changing corner position towards the middle.

Furthermore, provided the pick-up and bow position are closer to the bridge than to the nut or the finger pressing the string (which is the case for the majority of finger positions), the longest of either the measured up or down going oscillation sections will correspond to the case where the Helmholtz corner is traveling on the nut (or finger) side of the string and where the string is sticking to the bow (as indicated in figure 1). Since in this state the string and the bow nearly coincide in velocity [9], the up or down movement reveals the bowing direction (and an approximation of the bow velocity). Therefore, the relative corner position is expected to alternate between the $[0, 0.5]$ and $[0.5, 1]$ ranges when bow direction changes occur.

3.3.3 Normalised corner position

Another physical fact related to the corner position can be derived from the theory in section 2. In an ideal Helmholtz motion the relative corner position matches the relative pick-up position in the freely vibrating part of the string (which can be estimated from the priorly measured pick-up position and total string length, and using the estimated fundamental frequency). Hence, it can be expected that the measured deviation from that theoretical prediction is dictated solely by the mentioned interplay between acoustic losses and bow-string-interaction properties.

In other words, while fingered notes change the relative pickup-position and thereby the relative corner position, the ratio of those features, hereafter referred to as the "normalised corner position", should provide a feature whose deviation from 1 or -1 reveals a more note-independent indication of the player's input.

3.3.4 Root-mean-square deviation (RMSE)

The limitations of the model also potentially carry useful information, which can be studied by comparing the segmented signal and the original data. For now, we only consider the total root-mean-square deviation (RMSE) of this residual error, which can be interpreted as a lumped feature, comprising the combined effects of all other so far

[3] We note that the duty cycle of the signal reflects the stick and slip regions as they pass by the point of measurement, which would be slightly different than the actual stick and slip regions at the point of bowing.

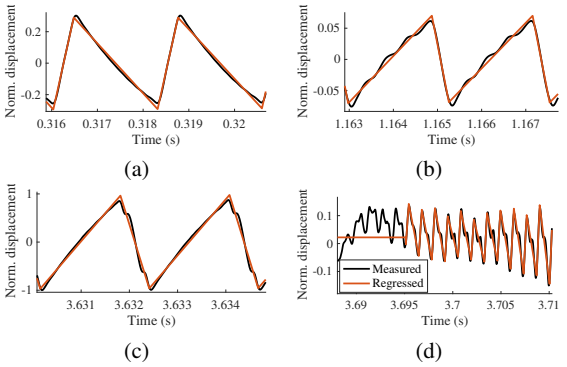

Figure 3. Four examples of segmented regressions of measured violin string displacement signals. The black and orange curves respectively show the measured and regressed signals.

neglected properties of the original signal. From a physical viewpoint, most deviations from the ideal Helmholtz motion are roughly proportional to the amplitude and period of the signal. Hence, it is more appropriate to express the error as relative to these features. Meanwhile, from a perceptual viewpoint, this independent expression of the error feature makes it more consistent with timbre.

4. RESULTS

We carried out an evaluation where we instructed an advanced violin player to perform a set of specific techniques and a series of musical excerpts. The separately measured string movement was recorded and analysed with the presented algorithm. By comparing the feature results with various amounts of segments N in the regression, it was found that a number of $N = 6$ segments results in the optimal balance between low-noise features and latency (a more detailed report on this is out of the scope of this paper), and convergence was obtained in about 10 to 30 iterations with $\gamma_{\text{th}} = 10^{-6}$.

It must be stressed that the purpose of this paper is not to solve a physical model to provide performance parameters such as speed or force of the bow, but rather to generate a parametrisation that responds meaningfully to the changes in the performer's actions, which can later be used in real-time performance applications. This means that there is no obvious ground truth to compare against in this case.

4.1 Segmented Regressions

Figure 3 shows four examples of segmented linear regressions of oscillation cycles. Figures 3 (a), (b) and (c) are all steady-state regimes of A4 notes, but they are played with different playing techniques, which can be noted from figure 4, where reference to these figures is made in the time axis. Meanwhile, (d) is an attack transient of an F#5 note.

In most cases, the regressed segmented signal is reasonably close to the measured signal. Especially the shorter segment, corresponding to the bow-string slip-phase, is generally very linear, except for rare cases such as in figure 3 (c) where "multiple flybacks" occur, i.e. the slip-state is interrupted by short intermediate sticking phases. It can be also noted that also all non-modelled effects in these examples repeat themselves exactly on consecutive cycles, i.e. they also represent deterministic steady-state components. All examples also show rounded corners, but it is worth mentioning that other signal examples showed a wider variety of corner roundings (not plotted). In figure 3 (d), the low amplitude at the beginning of the note onset proves vulnerable for bowing gestures causing an offset in the signal as the first oscillation cycles are not detected. It can be also seen, in this example, how a second slip (i.e. a second harmonic) occurs but fades out again quickly.

4.2 Sound Features

Figure 4 (a) shows the obtained features for a few particular playing techniques, which are annotated at the bottom [4]. We note that that the data used for this figure only contains samples without bow changes. Figure 4 (b) shows the features for the first twelve notes of J.S. Bach's II Double BWV 1002, along with indication of note-changes and the played note and string.

The fundamental frequency is compared with an f0 estimation provided by the Sonic Visualiser software, which relies on the LibXtract library [23, p. 69] that implements an estimation based on the "Average Magnitude Difference Function". It should be noted that this estimator uses an at least four times larger time frame than the oscillation periods, resulting in a low-pass filtered yet slower response.

4.2.1 Varying bow speed

It is interesting to note that the normalised corner position deviates more from unity for the lower bow speed. As can be seen in figure 3 (b), it is likely that the ripples in the sticking section are biasing the actual corner position. A closer examination confirmed that a virtually equal corner position is found between high and low bow speed when the feature is derived from the initial segmentation (i.e. using the minimum-to-maximum peaks), which may support this hypothesis. This example illustrates how effects that are not taken into account in the model can slightly influence the features. While there is no significant change in the average fundamental frequency with changing bow speed, there is a clearly increased noisiness of this feature in the low bow speed case. This may be related to the fact that low oscillation amplitudes are more sensitive to bowing gestures, which can also be perceived as a more hesitating sound.

4.2.2 Varying bow force

While a change in bow force also evokes a change in amplitude, it does not affect the corner position so much, but rather the fundamental frequency and its noisiness seems to show a noticeable correlation.

4.2.3 String release

As predicted by the theory, when the bow is released from the string, the corner position converges towards 0.5. This

[4] It should be noted that, in this brief study, the evaluation of the bowing parameters hasn't been strictly independently evaluated.

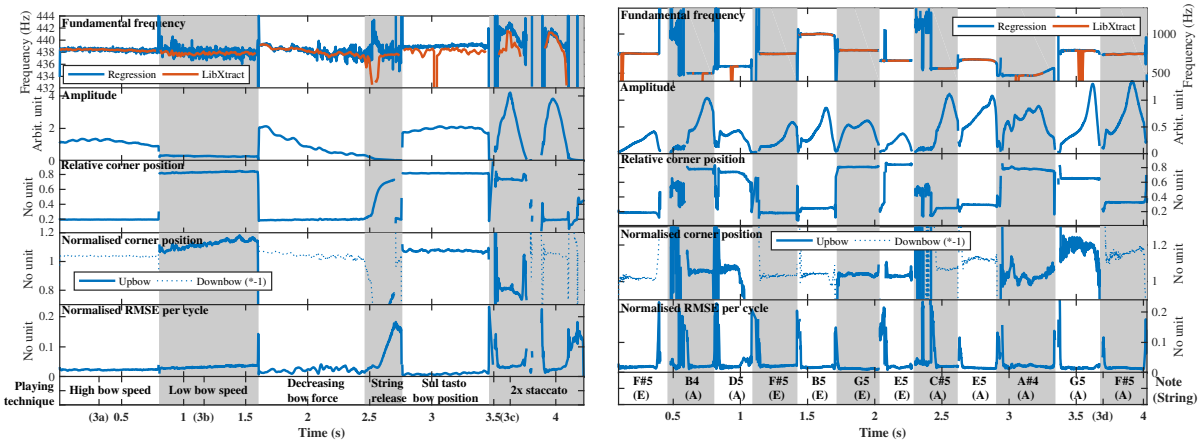

Figure 4. Obtained features for signal samples of various playing techniques (a) and of a musical phrase (b). Playing techniques and played notes are indicated at the bottom rows. The labels (3a), (3b), (3c), (3d) in the timeline refer to the respective plots of figure 3.

appears to occur quite fast, and it is also worth noting that it even surpasses this value. The increase in RMSE can be explained by the fact that the gradually rounding wave shape diverges more and more from the triangular Helmholtz shape. However, the fastest identification of this particular regime seems to be provided by the corner position features. The more precise representation of the normalised corner position reveals a clearly abrupt transition when the bow releases the string.

4.2.4 *Sul tasto bow position* [5]

This bowing technique results in a very good approximation of the Helmholtz motion, which can be noted from the remarkably low RMSE feature. Because of this low RMSE, we can confidently say that the fundamental frequency obtained from the regression is not corrupted by falsely detected cycles, and that it are failures of the LibXtract method that are at the cause of the deviations between the curves.

4.2.5 *Staccato*

As shown in figure 3 (c), the first of the played staccato notes is marked by multiple flybacks in the slip regime. Meanwhile, the wave shape in the second staccato note is similar to 3 (a). The flyback regime may well be the cause of the former's unusually low normalised corner position.

4.2.6 *Musical phrase*

As theoretically predicted, most bow direction changes can be clearly identified from the swap in corner position. It would be useful if there were a feature threshold that enables separation of note onsets and more steady-state oscillation regimes. With the current model, the note onsets cause erroneous initial segmentations which lead to the high RMSE values noted in figures 3 (b) and 3 (d), which would therefore be a potential candidate to identify note-onsets. However, closer examination revealed that there is no RMSE threshold that enables fully consistent identification.

Since various finger positions are used in this excerpt, it can be noted that the normalised corner position indeed confirms a behaviour that is more independent from the note than the relative corner position on its own.

While there appears to be an erroneous leap in pitch at the end of the first note, closer observation reveals an ambiguous fundamental frequency situation due to a strong double-slip oscillation regime. It can be noted that many dropouts occur in the fundamental frequency obtained by LibXtract, which is mainly due to the low performance of this method at low amplitudes.

It is further interesting to note that the notes E5, F#5 and G5 are played both on the E and A string. The normalised corner position for all of these notes is somewhat higher when played on the A string (unlike the low bow speed note in figure 3 (a), there are no ripples on the signal, excluding this potential cause). While a more comprehensive study will be needed to draw pertinent conclusions, it can be hypothesised that there is an inverse correlation with the (vibrating) string length, which in turn may be related to the amount of acoustic losses.

5. CONCLUDING REMARKS

Real measured string displacement signals appear to behave reasonably close to the ideal Helmholtz motion of string vibration. Therefore, we hypothesise that approximating the measured signal by this signal model using regression yields meaningful parameters with regard to the character of the sound and therefore to the player's musical intent.

As a next step, the extracted parameters take a form that could easily be adapted to control a musical synthesiser. The frequency and amplitude parameters have obvious mappings for this application. While these can also be extracted with more generic signal-processing algorithms (e.g.

[5] Sul tasto means bowing near or over the fingerboard.

[7]), our system takes advantage of the known characteristics of the bowed string. This specific-purpose approach can provide less ambiguous features regarding the physics taking place, with low and constant latency (compared to spectral or windowed methods); which we believe can significantly improve fast and reliable quantisation of detailed musical intentions. Our approach also allows a tuning between latency and precision based on the number of segments to be fit at a time.

The corner position enables detection of bow direction, which could be used to shape the output sound, and its converge toward 0.5 when the bow leaves the string could change the state of a connected synthesiser from sustain to release, even before the violin sound entirely stops. Meanwhile, the deviation of the normalised corner position from ± 1 appears to be an indicator of the string length (yet it may be biased by non-Helmholtz signal deviations, which will be considered later). The periods of increased RMSE often correspond to bow changes in a way that could detect note onsets with less latency than the spectral methods that are often used for onset detection on string instruments. Meanwhile, low RMSE values may enable the identification of the sul tasto playing style. That being said, it should be stressed that the goal of this work is not to extract physical parameters of violin performance, but it suggests a relationship with timbre or tone quality.

Finally, the systematic deviations identified in Figure 3 between the original signal and the segmented linear regression point the way toward further, more detailed feature extraction which could be used for performance applications.

The purpose of this paper has been to lay out the theoretical and mathematical foundations of a real-time feature extraction algorithm which draws on the physics of the bowed string. The next stages of this work will apply the results to real-time sound synthesis and consider the subjective response of the performer to the resulting digital instruments.

Acknowledgments

This work is supported by EPSRC under grant EP/N005112/1.

6. REFERENCES

[1] T. Jehan and B. Schoner, "An audio-driven perceptually meaningful timbre synthesizer," in *In. Proc. Int. Comp. Music Conf.*, 2001, p. 8.

[2] G. Essl and S. O'Modhrain, "An enactive approach to the design of new tangible musical instruments," *Organised Sound*, vol. 11, no. 03, p. 285, 2006.

[3] J. Janer and E. Maestre, "Mapping phonetic features for voice-driven sound synthesis," *Commun. Comput. Inf. Sci.*, vol. 23 CCIS, pp. 304–314, 2008.

[4] P. A. Tremblay and D. Schwarz, "Surfing the Waves : Live Audio Mosaicing of an Electric Bass Performance as a Corpus Browsing Interface," *Proc. 2010 Int. Conf. New Interfaces Music. Expr. (NIME10), Sydney, Aust.*, pp. 447–450, 2010.

[5] M. Puckette, "Grafting Synthesis Patches onto Live Musical Instruments," *Proc. Int. Comput. Music Conf.*, vol. 2010, pp. 250–253, 2010.

[6] K. McMillen and C. Shaver, "Computer interface for polyphonic stringed instruments," feb 2010.

[7] C. Poepel, "An investigation of audio signal-driven sound synthesis with a focus on its use for bowed stringed synthesisers," Ph.D. dissertation, University of Bermingham, 2011.

[8] M. M. Wanderley and P. Depalle, "Gestural control of sound synthesis," *Proc. IEEE*, vol. 92, no. 4, pp. 632–644, 2004.

[9] J. Woodhouse and P. M. Galluzzo, "The Bowed String As We Know It Today," *Acta Acust. united with Acust.*, vol. 90, no. May, pp. 579–589, 2004.

[10] A. Perez-Carrillo, "Statistical models for the indirect acquisition of violin bowing controls from audio analysis," in *Proc. Meet. Acoust. 172ASA*, vol. 29, no. 1. ASA, 2016, p. 35003.

[11] P. R. Cook, "Physically informed sonic modeling (phism): Synthesis of percussive sounds," *Comput. Music J.*, vol. 21, no. 3, pp. 38–49, 1997.

[12] A. Farnell, "An introduction to procedural audio and its application in computer games," *Audio Most. Conf.*, no. September, pp. 1–31, 2007.

[13] L. Nijs, M. Lesaffre, and M. Leman, "The Musical Instrument as a Natural Extension of the Musician," *Castellengo, M. Genevois, H. (Eds.). Music its instruments. Sampzon Ed. Delatour Fr.*, pp. 132–133, 2013.

[14] M. Leman, M. Lesaffre, L. Nijs, and A. Deweppe, "User-oriented studies in embodied music cognition research," *Music. Sci.*, vol. 14, no. 2_suppl, pp. 203–223, 2010.

[15] H. von Helmholtz, "Lehre von den Tonempfindungen. Braunschweig, 1862. English edition: On the sensations of tone," 1954.

[16] C. V. Raman, "On the mechanical theory of the vibrations of bowed strings and of musical instruments of the violin family, with experimental verification of the results," *Indian Assoc. Cultiv. Sci. Bull*, vol. 15, pp. 1–158, 1918.

[17] G. Weinreich and R. Caussé, "Elementary stability considerations for bowed-string motion," *J. Acoust. Soc. Amer.*, vol. 89, no. 2, pp. 887–895, 1991.

[18] L. Cremer, *The Physics of the Violin*. Cambridge, MA: MIT Press, 1984.

[19] M. E. McIntyre, R. T. Schumacher, and J. Woodhouse, "On the oscillations of musical instruments," *J. Acoust. Soc. Amer.*, vol. 74, no. 5, pp. 1325–1345, 1983.

[20] V. M. R. Muggeo, "Estimating regression models with unknown break-points," *Stat. Med.*, vol. 22, no. 19, pp. 3055–3071, 2003.

[21] G. E. P. Box and P. W. Tidwell, "Transformation of the Independent Variables," *Technometrics*, vol. 4, no. 4, pp. 531–550, 1962.

[22] V. M. R. Muggeo, "Segmented: An R package to Fit Regression Models with Broken-Line Relationships," *R News*, vol. 8, no. May, pp. 20–25, 2008.

[23] J. Bullock, "Implementing audio feature extraction in live electronic music," 2008.