SCORE LEVEL TIMBRE TRANSFORMATIONS OF VIOLIN SOUNDS

Alfonso Perez, Jordi Bonada, Esteban Maestre, Enric Guaus, Merlijn Blaauw

Music Technology Group,
Universitat Pompeu Fabra,
Barcelona, Spain
{aperez,jbonada,emaestre,eguaus,mblaauw}@iua.upf.edu

ABSTRACT

The ability of a sound synthesizer to provide realistic sounds depends to a great extent on the availability of expressive controls.

One of the most important expressive features a user of the synthesizer would desire to have control of, is timbre. Timbre is a complex concept related to many musical indications in a score such as dynamics, accents, hand position, string played, or even indications referring timbre itself.

Musical indications are in turn related to low level performance controls such as bow velocity or bow force. With the help of a data acquisition system able to record sound synchronized to performance controls and aligned to the performed score and by means of statistical analysis, we are able to model the interrelations among sound (timbre), controls and musical score indications.

In this paper we present a procedure for score-controlled timbre transformations of violin sounds within a sample based synthesizer. Given a sound sample and its trajectory of performance controls: 1) a transformation of the controls trajectory is carried out according to the score indications, 2) a new timbre corresponding to the transformed trajectory is predicted by means of a timbre model that relates timbre with performance controls and 3) the timbre of the original sound is transformed by applying a time-varying filter calculated frame by frame as the difference of the original and predicted envelopes.

1. INTRODUCTION

In this work we are presenting timbre transformations within a violin synthesizer based on samples, providing the user with the possibility of controlling the timbre of the produced sound. Different annotations directly related to timbre are used in musical scores, including dynamics (*p, mf, f*), specific string and hand position and bow-bridge distance (sul tasto, ponticello). Non standard annotations can also be provided, for example predefined labels indicating a specific timbre: sweet, hard, brilliant, etc.

Apart from dynamics, we can clearly perceive two different timbre regions in violin playing: A brilliant timbre, typical for notes played close to the bridge and at high forces, and a soft timbre typical for notes played further from the bridge at lower forces. Between these two extremes we find a wide range of different timbres.

By means of a 3D motion tracking system ([1], [2]) we are able to obtain performance controls carried out by the violinist. Motion data is recorded synchronously with sound and aligned with the musical score being performed. This way we build a database of sound and control data labeled with musical annotations. By means of machine learning techniques we are able to find relations among the three domains and we can deduce how controls are influencing timbre [3].

The main controls considered are bow transversal position, bow velocity (derivative of bow position), bow acceleration (second derivative of bow position), bow force, string being played, finger position (distance to the bridge), bow-bridge distance and β (bow-bridge distance relative to effective length of the string - given by finger position). Timbre is defined as the spectral envelope calculated as the energy in 40 frequency bands in a logarithmic scale.

It is also important to consider the interrelations among control parameters. From the whole control space only some parts are of interest in a traditional musical sense. Only those regions in which the vibration of the string is in Helmholz motion regime [4]. This is referred to as the *playable space* in the literature and it is the essence of the known Schelleng's diagram [5]. The diagram shows the boundaries of the playable space relating force and β for a given velocity. Interrelations among parameters implies that a change in one parameter will affect all the rest.

In this paper we will present timbre transformations based on user musical indications. Given a (source) sound, its performance controls , and a user annotation, this transformations are carried out in three steps: 1) Source control parameters trajectory is adapted to the user musical annotation, i.e. a B4 note sample played on the A-string, has to be transformed as if played on the D-string. This means a control parameter change (string played) that involves finger position change (B4 note on D-string is obtained at about 20 cm from the bridge and on A-string at about 34 cm). It affects β because the effective string length is shortened and this involves in turn a change in bow force and bow velocity (according to Schelleng); 2) Timbre (spectral envelope) is predicted for the adapted control trajectory by means of a timbre model that relates controls to timbre; and 3) A time-varying filter is calculated as the spectral envelope difference of the source sound and the predicted one, frame by frame. The filter is applied to the source sound. Here we are assuming that the timbre of a sound can be transformed into another by applying the difference filter.

The transforming procedure is presented along the paper in a reverse way in order to validate each of the enumerated steps: in Section 2 we describe the filtering stage and make sure that it is possible to transform the timbre of a sample by applying a filter calculated as the difference of source and target envelopes. In Section 3 we present the timbre model that is used to predict the spectral envelopes and in Section 4 we visualize the playable space for the main control parameters and we propose a method to drag control trajectories from one part of the space to another.

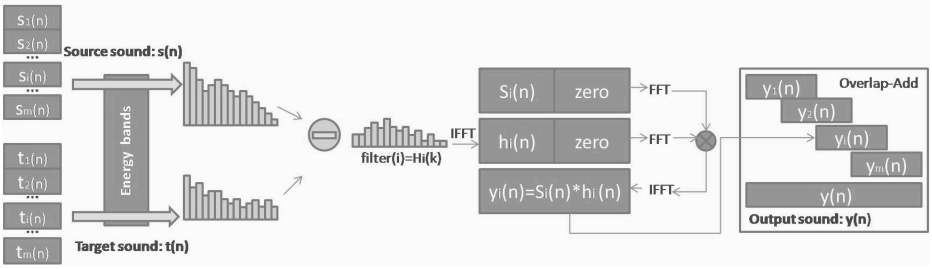


Figure 1: *Obtaining and applying the filter*

2. OBTAINING THE FILTER

This section presents how the varying filter is obtained and applied. The idea is to divide the source sound into frames and for each frame calculate its spectral envelope and filter it accordingly to a target frame spectrum. As a proof of concept and in order to validate the filtering procedure we can transform a source sound by using the spectral envelope of a recorded target sound and verify by listening if target and transformed sounds are perceptually similar.

With this purpose, several notes with the same pitch (A4, 440Hz) and duration (4 seconds) but performed in different manners, were recorded: 1) A-string, piano, far from the bridge; 2) A-string, mezzoforte close to the bridge; 3) D-string (4th position), piano, far from the bridge; 4) D-string, mezzoforte, close to the bridge.

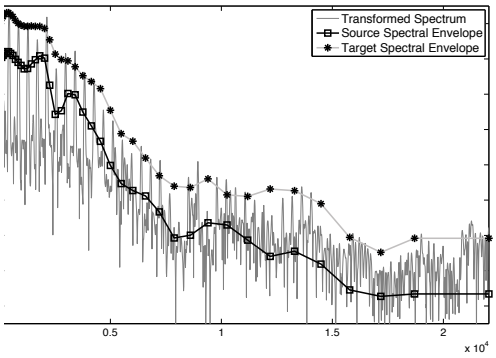
The selection of these different ways to play the A4 note is done based on looking for the biggest perceptual differences among them. Playing far from the bridge or at high hand positions, we obtain a softer sound, whereas playing close to the bridge and at lower positions we obtain a more brighter tone. The notes were timbrically transformed, ones into the others, and their similarity after the transformation was compared.

The transformation procedure (Figure 1) consists of calculating the spectral envelope of each frame for two notes (source and target). Then calculate the filter as the difference of the envelopes at each frame, smooth the filters in time to avoid frame to frame discontinuities and apply the obtained smoothed varying-filter to the source note as in [6]. The transformed note is finally compared with the target note by listening.

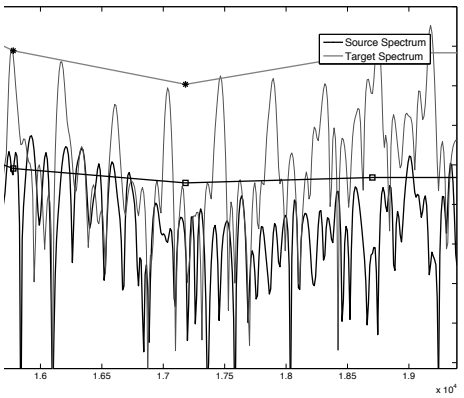
Spectral envelope is calculated as the harmonics energy in 40 overlapped frequency bands distributed along the frequency axes in a logarithmic scale.

By listening to the transformations we can conclude that although there are some differences between transformed and target sounds they are very similar regarding timbre and it could give the impression that they were played using the same control parameters.

In Figure 2(a), the spectrum envelope of a frame of an A4 note played on the A-string far from the bridge (source note) is plotted as a dark line with squared markers; the light line with asterisks corresponds to the spectrum envelope of a note played close to the bridge (target note); The transformed spectrum is plotted in an intermediate tonality. We can notice how the transformed spectrum fits to the target envelope. In Figure 2(b) there is a zoom of the high frequencies for the spectra of both source and target notes. We can realize how notes played closer to the bridge (light with asterisks) have more stiff harmonics and thus less noise. When applying the filter to the source sound, non-harmonic components are also amplified. This means that the filter has to be applied only to the harmonics. An extra energy envelope filter for the non-harmonic components may produce a more realistic sound.



(a) Source and Target Spectral Envelopes with transformed spectrum



(b) Source and Target Spectra and envelopes. High frequencies

Figure 2: *Spectral Envelope Transformation*

3. PREDICTING SPECTRAL ENVELOPES

Assuming that we can transform samples by using the spectral envelope of the target sample, the next step is to predict spectral envelopes given the sequence of performance controls of the target sound.

Here is where it comes out the timbre model that relates con-

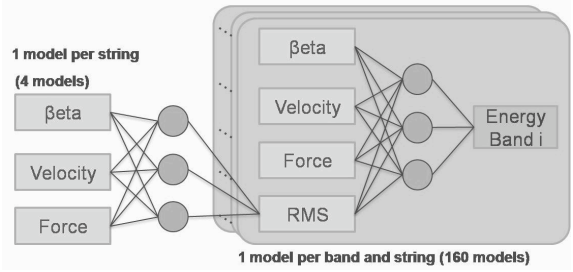


Figure 3: *Typical network architecture for a RMS-relative model with three input control parameters*

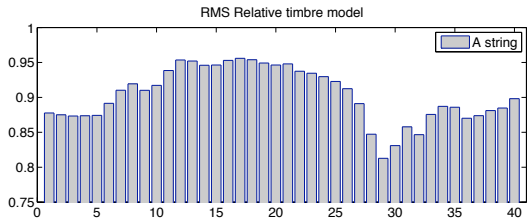


Figure 4: *Correlation coefficient for each band.*

trol parameters (bow velocity, bow force etc.) to the corresponding spectral envelope (calculated as energy in frequency bands as explained previously). As described in our previous work [3], the timbre model is based in neural networks. In this section we briefly present an extension to it where we include the use of temporal information and we predict the energy of the bands relative to the Root Mean Square Energy (RMS from now on) in the whole frame.

Different combinations of inputs and outputs for the networks were tested. The controls that seem to influence timbre the most are β, bow force and bow velocity, but others such as bow tilt or bow position seem to be also important. Prediction in sustained parts of the sound is very accurate. During transients (specially in note attacks) temporal information helps to improve the prediction. As output we are predicting the energy in the frequency bands. This means that we have a neural network per string predicting the frame-energy at each band (40 bands * 4 strings = 160 networks).

Temporal information can be introduced by adding as inputs the derivatives of the control parameters and by feeding the input not only with performance controls of the current frame but also of previous frames.

These models would predict energy in a band independently of the others, but they are actually correlated. In a frame with low energy, all bands will have low energy and the envelope should be smooth without big jumps between consecutive bands. In order to make the bands somehow dependent among them, we propose to predict the energy of each band relative to the RMS energy of the actual frame, and therefore, we also need to predict the RMS energy in a previous step. The typical architecture of these RMS-relative model networks is as depicted in Figure 3 for three input parameters. With such an architecture errors are quite constant, independently of the band.

Performance of the models depends on several factors such as the learning dataset, the type of model, the band, the inputs to the model, the string or the type of sound (sustained or transient). In the case of a RMS-relative model such as the one in Figure 3, averaging errors for all bands and strings we get a correlation coefficient between real and predicted energy bands of around 0.879 (with a ten-fold cross-validation). In Figure 4 we can observe the correlation factor for each band on the A-string. The coefficient is quite constant among the bands, varying from around 0.85 to 0.95.

4. PERFORMANCE TRAJECTORY TRANSFORMATION

Provided that we are able to predict the timbre of a sample given its trajectory of performance controls (Section 3) and that we can transform the timbre of the sample to match the predicted one (Section 2), it only remains how to obtain the target controls given some user indications in the score. Here we are proposing to transform the controls trajectory to fit inside another part of the control space. The complete schema for timbre transformations is depicted in Figure 5.

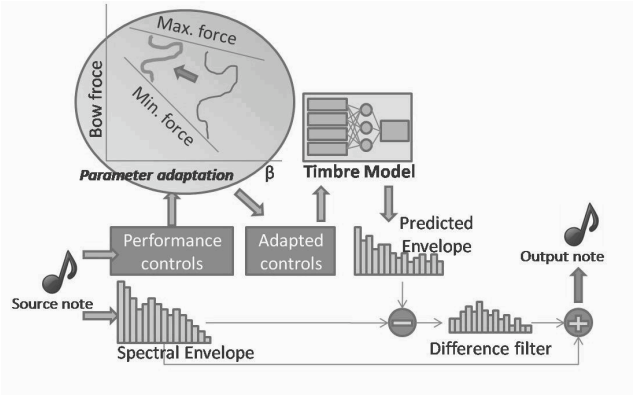


Figure 5: *Schema for timbre transformation: a)shift and adapt controls, b)predict new timbre, and c)filter*

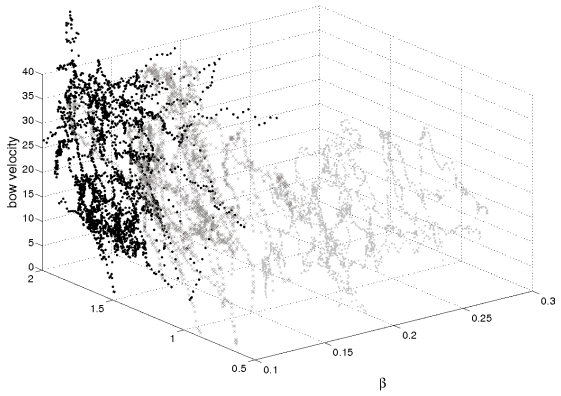
As an example of score driven transformations we propose the followings:

- Dynamics change, among a set of predefined categories, i.e.:*piano, mezzoforte, forte)*.
- String and hand position change: same pitch note played in a different string.
- Timbre change, among a set of predefined categories, i.e.: *brilliant, soft*.

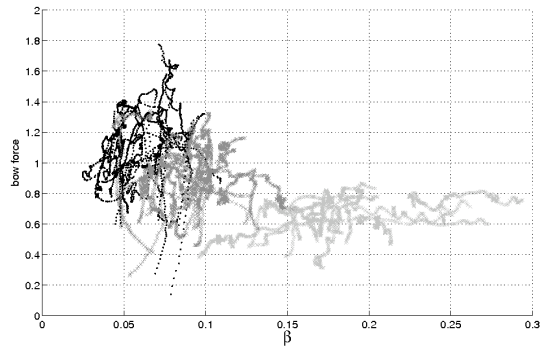
All these transformations are directly affecting some control parameters: string played, β, etc. Control parameters are very interrelated. When changing one, the others need also to be adjusted in order to obtain a musically acceptable sound keeping the string vibrating within the Helmholtz motion regime. For example, at a constant bow velocity, if we move closer to the bridge, the range of possible bow force is smaller and maximum and minimum force boundaries are increased. The combination of parameter values

out of which we can't obtain a Helmholz vibration, defines the playable space. The known Schelleng's diagram [5] represents this playable space in two dimensions for slow bow velocities (from 10 to 20 cm/s). With the help of the data acquisition system described in [1] we can obtain control data of real performances, so we can cover the entire range of possible control values and we can extend the playable space model to include many other parameters.

A 3D visualization of the control space for the parameters force, β and velocity is represented in Figure 6. It was obtained with a small database of 47 notes played on the A-string. A more complete diagram can be obtained by including more samples and by forcing the performer to play at boundary regions. Color tonalities represent different dynamics. Samples with same dynamic tend to form clusters. This way we are including in the playable space not only control parameters but also perceptual ones. A transformation in the dynamics would require to shift the control trajectory from one cluster to another. Notice that bow pressing force is not given in Newtons, it is a measure of distance as explained in [1], so diagrams can not be contrasted with Schelleng's.



(a) *3D distribution*



(b) *β vs force*

Figure 6: *Distribution of parameters β*, bow force *and* bow velocity *in 3D and β-vs-force projection for A-string. Tonalities correspond to dynamics: piano (light), mezzoforte (medium), forte (dark).*

A very preliminary method to transform control trajectories is proposed: 1) Set the values for the parameters directly related to the score annotation (i.e.: for a string change, string and finger position are implicitly given). 2) Look weather each of the rest of the parameters is inside the playable subregion given the change in the parameters in 1 (i.e: A sample played *forte* with bow force values in the range of [1, 2]cm, is transformed to *piano*. Forces are then out of the *piano* region, so values must be shifted to the range [0.3 to 0.8]cm). 3) If the range of values of the target region is smaller than for the source values, then values have to be scaled to fit the new range.

5. CONCLUSIONS

We presented a timbre transformation procedure for recorded violin samples based on performance controls where timbre transformations are driven by indications in a score. The main elements involved in the transformation are a timbre model that relates control data with timbre and a control trajectory transformation procedure that shifts and scales control trajectories inside a playable region.

This is a work in progress that needs some improvements, among them, a second spectral envelope model for non-harmonic components of the sound, a more complete model of the playable space by including more samples and control parameters, and an evaluation of the whole procedure that until now has only been tested with isolated samples.

6. ACKNOWLEDGMENTS

This work has been supported by Yamaha Corp.

7. REFERENCES

[1] Esteban Maestre, Jordi Bonada, Merlijn Blaauw, Alfonso Perez, and Enric Guaus, "Acquisition of violin instrumental gestures using a commercial emf device," in *Proceedings of International Computer Music Conference*, Copenhagen, Denmark, 2007.

[2] Enric Guaus, Jordi Bonada, Alfonso Perez, Esteban Maestre, and Merlijn Blaauw, "Measuring the bow pressing force in a real violin performance," in *Proceedings of International Symposium on Musical Acoustics*, Barcelona, Spain, 2007.

[3] Alfonso Perez, Jordi Bonada, Enric Guaus, Esteban Maestre, and Merlijn Blaauw, "Combining performance actions with spectral models for violin sound transformation'," in *Proceedings of 19th International Congress on Acoustics*, Madrid, Spain, 2007.

[4] Lothar Cremer, *Physics of the Violin*, The MIT Press, November 1984.

[5] John C. Schelleng, "The bowed string and the player," *JASA*, vol. 53, pp. 26–41, 1973.

[6] Udo Zölzer, Ed., *Digital Audio Effects*, Wiley, 2002.