ABSTRACT

Though musicians tend to agree on the importance of practicing expressivity in performance, not many tools and techniques are available for the task. A machine learning model is proposed for predicting bowing velocity during performances of violin pieces. Our aim is to provide feedback to violin students in a technology-enhanced learning setting. Predictions are generated for musical phrases in a score by matching them to melodically and rhythmically similar phrases in performances by experts and adapting the bow velocity curve measured in the experts’ performance. Results show that mean error in velocity predictions and bowing direction classification accuracy outperform our baseline when reference phrases similar to the predicted ones are available.

KEYWORDS

expressive performance modeling, machine learning, violin, music education

1 INTRODUCTION

Becoming a skilled music performer requires a huge amount of practice, much of which is typically unsupervised. Music students regard that practicing expressive performance is particularly desirable, and that it tends to be neglected [3]. The increasing availability of technological resources for learning may offer an opportunity to bridge that gap. One way in which this can be done is by developing systems that provide real-time feedback about students’ performance. The CyberViolin [7] system uses electromagnetic motion tracking to classify in real-time the type of articulations played by a musician. In their envisioned application scenario, the correct articulation is known for each note, and the virtual tutor is in charge of verifying if students articulate the proper technique in each case. Similarly, Rasamimanana et al. [8] develop models to characterize the different articulation techniques in terms of bow velocity and acceleration. The MusicJacket [11] system explores new forms of providing performance training assistance in an experiment where students improve their bowing skills assisted by a wearable device with vibrotactile feedback capabilities. However, in any case like the above, the system must have a prior knowledge of a ground truth, that is, a reference to form the basis of its feedback.

In this work, we propose an approach to modeling bow velocity in violin performances based on score information, with the purpose of powering an interactive tutoring system so that students could be given suggestions and receive feedback about their bowing not only during performance of predetermined exercises, but when playing any piece, even if no reference performance of it is available.

2 BACKGROUND

Many models of expressive music performance have been proposed in the past [2]. The majority of them, however, target performance rendering by computers as described in [1], thus modeling sound features directly and failing to provide information on how to produce such sounds, i.e.: the motion of musicians.

A few studies modeling motion in violin playing warrant a review. Shan et al. [9] use motion capture data to identify physical markers that characterize the spiccato bowing technique, thus, the end goal is a model of performance, albeit one that is inferred by the researchers and expressed in terms of rules. The most relevant work in terms of modeling complex violin performance gestures from data is by Maestre [4], who derives bowing motion curves for combinations of note features (e.g.: détaché, bow down, pianissimo) with the purpose of improving violin synthesis. Our model, though simpler, aims to do the same in scenarios where fewer note features are available — essentially only pitches and durations — relying on the existence of similar training examples, much like a musician would choose to interpret a melody based on their memory of a similar one.

3 MATERIALS AND METHODS

The recordings used were made as part of experiments on ensemble expressive performance [5, 6]. The audio signals were recorded from a piezoelectric pickup attached to the bridge of the violin. Bowing motion data were acquired by means of a Polhemus Liberty wired motion capture system as detailed in Maestre [4], at a sample rate of 240 Hz. Synchronization of audio and motion capture data was reviewed through the alignment of linear timecode timestamps.
The general approach proposed for prediction is depicted in figure 1. Training set recordings are split into two–measure phrases. For each two–measure phrase in a new score, the closest match from the training set is determined and its bow velocity curve applied to the new phrase following the process indicated in figure 2. The measure of similarity between two phrases is given by applying the dynamic time warping algorithm to phrases as implemented by Stammen and Pennycook [10], so that it takes pitch contour and note duration ratios into consideration. The best match is assumed to be the phrase with the lowest warping cost. Therefore, the underlying assumption is that phrases which are melodically and rhythmically similar should be performed similarly.

Mainly two aspects of performance are reflected in measurements of bow velocity: the velocity modulus indicates how fast the bow was moving, which translates mostly into performance dynamics, and the velocity measurement sign is indicative of bowing direction. We follow slightly different approaches for predicting each of those: for bowing direction, after following the described process, each note is classified as “bow up” or “bow down” according to its predicted velocity sign. For velocity modulus prediction, the reference curve associated with phrases in the training set is the absolute value of its measured velocity.

4 RESULTS

A preliminary evaluation of the method was conducted following a leave–one–phrase–out procedure on a collection of phrases taken from the first violin in a performance of the fourth movement of Beethoven’s String Quartet no. 4, Op. 18. Motifs which repeated themselves throughout the piece were included only once to avoid positively biasing the nearest–neighbor algorithm.

4.1 Bowing Direction Prediction

A total of 68.75% of notes had their bowing direction correctly classified. In contrast, the baseline most frequent label classifier shows an accuracy of 50.94%.

Though it seems possible to improve accuracy with simple modifications to this model, we consider bowing direction classification as merely a quality indicator for the method, since our focus is on bowing velocity modulus prediction, which is more valuable due to its implications for expressive performance teaching.

4.2 Bowing Velocity Modulus Prediction

Figure 3 shows the distribution of absolute errors per note in predictions as a percentage of the full–scale value. The baseline model is the prediction of constant velocity equal to the mean absolute velocity value. A one–sided t–test indicates we are unable to reject the null hypothesis for equality of baseline and prediction means. Since this fact could be due to scarcity of data, we evaluate whether our measure of phrase similarity can be used as a predictor for bow velocity by splitting the predicted phrases in half, separating phrases with closest nearest neighbors from phrases with farthest ones. The bottom plot in figure 3 shows the distribution of errors for these two classes. In this case, t–testing confirms that phrases with closest neighbors present the lower mean absolute error of the two sets with $p = 0.01$.

Closer inspection of the data also indicates that even though nearest–neighbor approximations aren’t always successful in modeling note–level variations of absolute bow velocity, the mean predicted velocity of the phrase approximates the performed mean, as shown in figure 4. Since the absolute bow velocity curve and the loudness curve extracted from the audio recording share a high

Figure 1: Overview of the prediction method.

Figure 2: A graphical description of reference curve interpolation.
coefficient of determination ($R^2 = 0.76$), this could be interpreted as evidence that the melodic content of a phrase is indicative of its character in dynamics, i.e.: if a given melody is typically played *forte*, another, similar melody, should be played *forte* as well.

5 CONCLUSIONS

We have proposed a model for predicting bowing velocity in violin performances which is suitable for application in tutoring systems for giving suggestions and feedback to students. Such an application could make use of the predictions by, for instance, displaying a curve of the predicted velocity next to the score itself, or even color-coding the score based on that information. With these suggestions, novice students could be sensitized to the necessity of minding their expressive intentions, and experienced students could be challenged by different perspectives on how to express a piece.

The preliminary evaluation of the proposed model validates the use of melodic similarity as a predictor for bowing velocity in a phrase, but signals that the quality of predictions is closely related to the availability of sufficient samples to match all types of melodies present in the target score. A desirable next step would be to test the model in a learning environment, to verify if its predictions are capable of fostering improvements in bowing technique.

ACKNOWLEDGEMENTS

This work has been partly sponsored by the Spanish TIN project TIMUL (TIN 2013-48152-C2-2-R), the European Union Horizon 2020 research and innovation programme under grant agreement No. 688269 (TELMI project), and the Spanish Ministry of Economy and Competitiveness under the Maria de Maeztu Units of Excellence Programme (MDM-2015-0502).

REFERENCES


