Learning Violinist's Expressive Trends

Miguel Molina-Solana MIGUELMOLINA@UGR.ES
Department of Computer Science and Artificial Intelligence, University of Granada, 18071 Granada, Spain

Josep Lluis Arcos ARCOS@IIIA.CSIC.ES
Artificial Intelligence Research Institute, IIIA, CSIC, Campus UAB, 08193 Barcelona, Spain

Emilia Gomez EGOMEZ@IUA.UPF.EDU
Music Technology Group, Pompeu Fabra University, 08003 Barcelona, Spain

Abstract

This paper presents a Trend-based model for identifying professional performers in commercial recordings. Trend-based models characterize performers by learning how melodic patterns are played. Reported results using 23 violinists show the high identification rates achieved with our model.

1. Introduction

Expressive performance analysis and representation is a key challenge in the sound and music computing area. Previous research has addressed expressive music performance using machine learning techniques. To cite some, in (Saunders et al., 2008) they represent pianists' performances as strings; in (Ramirez et al., 2007) they study how to measure performance aspects applying machine learning techniques; and in (Stamatatos & Widmer, 2005) a set of simple features for representing stylistic characteristics of piano music performers is proposed.

We focus on the task of identifying professional violinists from their commercial audio recordings. Our approach is based on the learning of the *Trend Models* that characterize each performer. Trend models capture expressive tendencies and are learnt from audio descriptors obtained by using state-of-the-art audio feature extraction tools.

The whole process is done in an automatic way, using only the recordings (scores are not used). Since commercial recordings are so heterogeneous, it is really difficult to exactly translate the audio to an accurate score representation. We deal with this problem by using a more abstract representation that the real notes, but still close to the melody (i.e. instead of focusing on the absolute notes, we focus on the melodic surface).

Specifically, we deal with this task by (1) using a higher-level abstraction of the automatic transcription focusing on the melodic contour (Grachten et al., 2005); (2) tagging melodic segments according to the Implication-Realization (IR) model (Narmour, 1992); and (3) characterizing the way melodic patterns are played as probabilistic distributions.

2. Trend-Based Modeling

A trend model characterizes, for a specific audio descriptor, the relationships a given performer is establishing among groups of neighbor musical events. For instance, the trend model for the energy descriptor will relate, qualitatively, the changes of energy for a given set of consecutive ascending notes. The main processes of the system are:

Feature Extraction and Segmentation The first process consists on extracting audio features from recordings using an existing tool (Camacho, 2007). Using fundamental frequency information, note boundaries are identified and melodic segmentation is performed. For each note we collect its pitch, duration and energy. We used the IR model by E. Narmour to perform melodic segmentation. Each segment is tagged with its IR pattern.

Learning Trend Models Trend models capture the way different audio descriptors change in the different IR patterns. A trend model is represented by a set of discrete probability distributions for a given audio descriptor.

To generate trend models for a particular performer and note descriptor, we use the sequences of values extracted from the notes identified in each segment. From these sequences, each value is compared with respect to the mean value of the fragment and is transformed into two qualitative values meaning 'the descriptor value is higher than the mean', and 'the value is lower than the mean'. In the current approach, since we are segmenting the melodies in groups of three notes and using 2 qualitative values, eight (2^3) different patterns may arise.

Next, a probability distribution per IR structure with these patterns is constructed by calculating the percentage of occurrence of each pattern. Thus, trend models capture statistical information of how a certain performer tends to play. Combining trend models from different audio descriptors, we are improving the characterization of each performer. Trend models for both duration and energy descriptors have been learnt.

Classifying new recordings We are using a nearest neighbor (NN) classifier to generate a ranked list of possible performers for a new input recording. When a new recording is presented to the system, the feature extraction process is performed and its trend model is created. This trend model is compared with trend models learnt in the previous stage acting as class patterns.

The distance between two trend models, is defined as a weighted sum of the distances between their respective IR patterns (i.e. their respective probability distributions).

3. Results

We work with Sonatas and Partitas for solo violin from J.S. Bach. We tested our system by performing experiments with commercial recordings from 23 different violinists and using three movements: Mov.2 of Partita 1, Mov.6 of Partita 1, and Mov.5 of Partita 3. Each experiment consisted in learning trend models with one movement and then testing them with another movement. Figure 1 reports the results achieved in the experiments. Mov.2 versus Mov.6 experiments demonstrate the performance of the system by using two movements from the same piece. Mov.6 versus Mov.5 shows the performance of the system by using two movements from different pieces.

In experiments using movements from the same piece, the correct performer was majority identified in the first half of the list, while in movements from different pieces, the most difficult scenario, the 90% of identification accuracy is overcame at position 15. We can also observe that a 50% of success is achieved using the five first candidates in any case (doubling the 22% of a random classifier).

These results show that the model is capable of learning performance patterns that are useful for distinguishing performers. The results are promising, especially comparing with a random classification where the success rate is clearly outperformed.

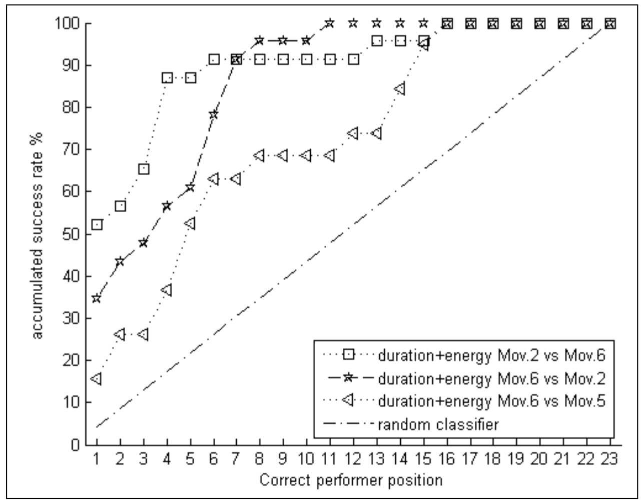

Figure 1. Accumulated success rate

We plan to increase the number of descriptors and to experiment with the rest of movements.

References

Camacho, A. (2007). *Swipe: A sawtooth waveform inspired pitch estimator for speech and music.* Doctoral dissertation, University of Florida, USA.

Grachten, M., Arcos, J. L., & Lopez de Mantaras, R. (2005). Melody retrieval using the implication/realization model. *MIREX 2005.*

Narmour, E. (1992). *The analysis and cognition of melodic complexity: The implication realization model.* Chicago, IL: Univ. Chicago Press.

Puiggros, M. (2007). Comparative analysis of expressivity in recorded violin performances. Master's thesis, Pompeu Fabra University.

Ramirez, R., Maestre, E., Pertusa, A., Gomez, E., & Serra, X. (2007). Performance-based interpreter identification in saxophone audio recordings. *IEEE Trans. on Circuits and Systems for Video Technology*, *17*, 356–364.

Saunders, C., Hardoon, D., Shawe-Taylor, J., & Widmer, G. (2008). Using string kernels to identify famous performers from their playing style. *Intelligent Data Analysis*, *12*.

Stamatatos, E., & Widmer, G. (2005). Automatic identification of music performers with learning ensembles. *Artif. Intell.*, *165*, 37–56.