Computational modelling of expressive music performance in hexaphonic guitar

Marc Siquier, Sergio Giraldo, and Rafael Ramirez

Universitat Pompeu Fabra (UPF), Barcelona, Spain
marc.siquier01@estudiant.upf.edu
sergio.giraldo@upf.edu
rafael.ramirez@upf.edu

Abstract. Computational modelling of expressive music performance has been widely studied in the past. While previous work in this area has been mainly focused on classical piano music, there has been very little work on guitar music, and such work has focused on monophonic guitar playing. In this work, we present a machine learning approach to automatically generate expressive performances from non expressive music scores for polyphonic guitar. We treated guitar as an hexaphonic instrument, obtaining a polyphonic transcription of performed musical pieces. Features were extracted from the scores and performance actions were calculated from the deviations of the score and the performance. Machine learning techniques were used to train computational models to predict the aforementioned performance actions. Qualitative and quantitative evaluations of the models and the predicted pieces were performed.

Keywords: Machine learning, Computational models, Expressive music performance, Hexaphonic guitar

1 Introduction

Computational modelling of expressive music performance deals with the study of the deviations from the score that musicians introduce when performing a musical piece (aka. *Performance Actions (PAs)*). Previous studies have mainly focused on monophonic piano classical music. Some exceptions include guitar and saxofone jazz music, in which only monophonic performances have been considered.

In this work we present an approach to computationally model polyphonic guitar performances from hexaphonic guitar recordings. The present approach is an extension of previous work on monophonic expressive performance on jazz guitar [1]. Hexaphonic guitar recordings of musical pieces recorded by a professional guitarist, were obtained using a Roland GK-3 divided pickup. A new set of features was defined aiming to capture not only the melodic (monophonic/horizontal) context of the score, but also the harmonic (polyphonic/vertical) context, depicting the harmonic progression or simultaneity between notes. Score alignment using *Dynamic Time Warping* (DTW) was performed to extract PAs defined as *Onset Deviation* and *Energy Ratio* Later, machine learning models were trained in order to predict the aforementioned PAs. Quantitavie evaluation of the models was performed by means of accuracy measures over both *train* and *cross-validation* schemes, as well as qualitative evaluation was assessed from perceptual tests of the synthesized predicted pieces.

2 Background

In the past, music expression has been mostly studied in the context of classical music, and most research focus on studying timing deviations (onset nuances [2]) and dynamics (energy [3]). There are several expert-based systems, such as the *director musices* [4] by the KTH group, studying this field from different perspectives. On the other hand, Machine-learning-based systems try to automatically obtain the set of rules to predict the PAs. For an overview of theses methods see Goebl 2005 [5]. Kirke et al [6] model polyphonic piano showing that multiple polyphonic expressive actions can be found in human expressive performances.

Previous work on guitar expressive performance modelling has been done by Giraldo et al. [1] and [7], who model ornamentation and PAs in monophonic jazz guitar performances, using machine learning techniques. Bantula [8] models expressive performance for a jazz ensemble of guitar and piano extracting features for chords such as *density, weight* or *range.*

3 Methodology

In Figure 1 we present a block diagram of the whole system, where four main stages are depicted: data acquisition (guitar recording), audio to MIDI transcription, feature extraction and model computation, and finally MIDI synthesis. Expressive hexaphonic guitar recordings were obtained using the Roland GK-3 divided pick-up, which is able to separate the sound from each string [9].

The main output of this first stage was a new dataset consisting of hexaphonic recordings recorded by a guitar player with different performance intentions. This dataset consists of 3 audio recordings (one recording of *Darn that dream* a jazz standard by Jimmy Van Heusen and Eddie De.Lange and two recordings of *Suite en la* a classical piece by Manuel M. Ponce.) resulting in a total of 1414 recorded notes and their corresponding music scores saved as XML files.

Transcription of each individual string was computed using the YIN [10] algorithm and envelope-based note segmentation. The transcription of each string was added into a single MIDI file by having each string in a different channel. After doing performance to score alignment with the original score and the transcription of the expressive guitar performance using Dynamic Time Warping (DTW), feature extraction and PAs computation was performed. As the player was told to follow strictly the score, we can assume there are no structural differences between the music scores and the expressive music performance, so DTW can be applied directly.

Feature extraction was performed following an approach in which each note was characterized by its *nominal*, *neighbouring*, and *contextual* properties. *Nominal* descriptors refer to the intrinsic or intra-note properties of score notes. *Neighbouring* descriptors or inter-note descriptors refer to the relations of the note with its neighbouring or simultaneous notes. *Contextual* descriptors refer to the context of the song in which the note appears in (e.g. chords, key, mode, etc). A total amount of 34 features where extracted for each score note, plus two PAs representing onset and energy deviation from each score note to its matching performance note.

Several machine learning such as K-Nearest Neighbours, Decision Trees, Supervector Machines and Artificial Neural Networks were applied to model *Onset Deviation* (difference in time between performance onset of a note and its corresponding onset in the score) and *Energy Ratio* (ratio between performance note energy and its corresponding energy in the score).

$$Onset\_dev_i = Onset\_perf_j - Onset\_score_i$$

$$Energy\_rat_i = \frac{Velocity\_perf_j}{Velocity\_score_i}$$

Also, feature selection has been computed and analyzed in order to retrieve the subset of descriptors that better predict the studied PAs.

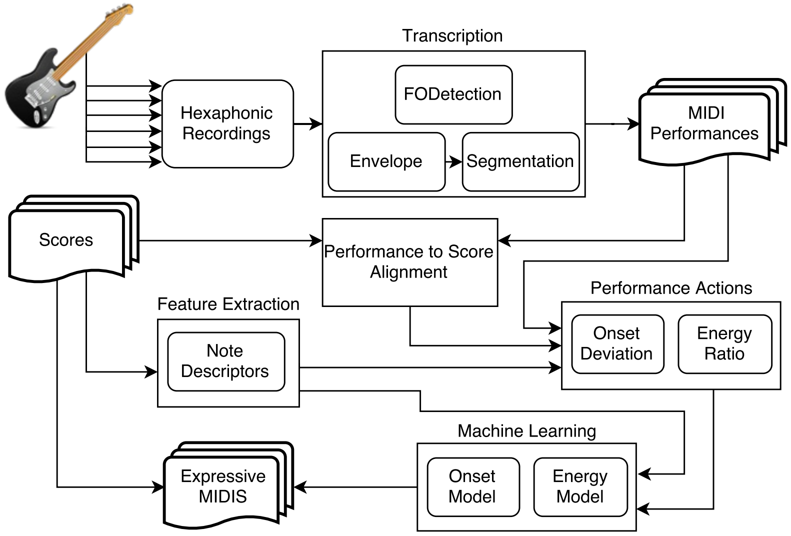

Fig. 1: Block diagram of the whole system.

4 Results

The proposed approach was quantitatively evaluated by measuring *Correlation Coefficient* (CC) obtained with the models studied, qualitatively evaluated by asking listeners to compare predicted and real performances. In Figure 2 we present the obtained CC for *Onset Deviation* and *Energy Ratio*. In red we show the accuracy for the whole training dataset and in blue the results applying 10-fold Cross-Validation (CV). The best accuracy (using CV) was obtained with the set containing the first 5 best ranked features. In Table 1 we show the results comparing different Machine Learning algorithms, both with CV and with the whole training set. We present the CC for *Energy Ratio* and *Onset Deviation* for the whole dataset, and using the best 5 features subset. The best results were achieved with Decision Trees where the obtained subset of 5 features outperforms the rest.

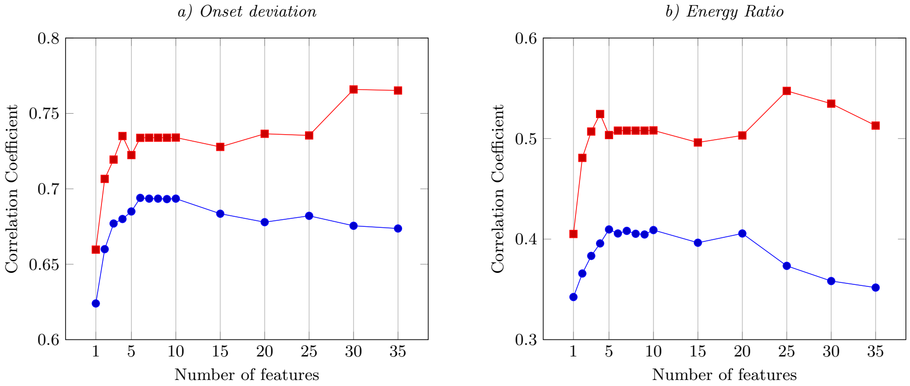

Fig. 2: Accuracies on increasing the number of features. Algorithm used: Decision Tree. Shown values correspond to Correlation Coefficients, in red for Train Dataset and in blue for 10 fold Cross-Validation.

Predicted feature	D.Tree cv/train	k_1NN cv/train	k_2NN cv/ train	SVM cv/train	ANN cv/train
Energy Ratio	0.35/0.51	0.22/1	0.26/0.78	0.21/0.33	0.23/0.63
Onset Deviation	0.67/0.77	0.30/1	0.36/0.81	0.39/0.45	0.29/0.67
Energy Ratio $_{5features}$	0.41/0.50	0.30/1	0.37/0.80	0.14/0.21	0.14/0.36
Onset Deviation $_{5features}$	0.69/0.72	0.38/1	0.61/0.82	0.30/0.31	0.44/0.43
Energy Ratio $_{bestsubset}$	0.41/0.51	0.30/1	0.37/0.79	0.16/0.21	0.15/0.39
Onset Deviation $_{bestsubset}$	0.69/0.73	0.37/1	0.58/0.82	0.30/0.32	0.48/0.48

Table 1: Results comparing different ML models (10 fold Cross-Validation). Shown values correspond to Correlation Coefficients.

For the qualitative survey, several synthesized pieces obtained by the models were compared to both the score (dead pan synthesis) and the performed (synthesized version) piece. Participants were asked to to guess how ”human” they sounded by comparing among them through an on-line survey [1]. Results from 15 participants (Figure 3) show that participants perceived the score more ”human” than the actual performance and predicted score. However, we obtained similar results among the performed piece and the predicted one, which might indicate that our models predictions are close to actual human performances.

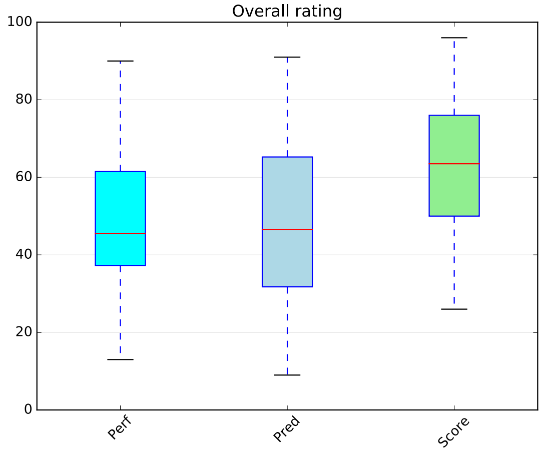

Fig. 3: Results of the on-line survey with performance, predicted and straight score synthesized midis.

5 Conclusions

In this work we have applied machine learning techniques in order to generate models for musical expression in polyphonic guitar music, by training different models for *Onset Deviation* and *Energy Ratio*. We treated polyphonic guitar as an hexaphonic instrument by capturing and transcribing each string separately. We extracted descriptors from the scores in terms of the melodic (Horizontal) as well from the harmonic (Vertical) context. We computed PAs from the aligned transcribed performance and the scores. We trained different models using machine learning techniques. Models were used to predict PAs that later were applied to the scores to be synthesized. Feature selection analysis and accuracy tests were performed to assess models performance. Perceptual tests were conducted on the predicted pieces to rate how close they sound to a human

[1] You can find the survey here: `https://marcsiq2.github.io/`

performance. Results indicate that descriptors contain sufficient information to generate our models able to predict performances close to human ones.

Acknowledgements

This work has been partly sponsored by the Spanish TIN project TIMUL (TIN2013-48152-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

1. Giraldo, S., Ramírez, R.: A machine learning approach to ornamentation modeling and synthesis in jazz guitar. Journal of Mathematics and Music **10**(2) (may 2016) 107–126
2. Sundberg, J., Friberg, A., Bresin, R.: Attempts to reproduce a pianist's expressive timing with Director Musices performance rules. Journal of New Music Research **32**(3) (2003) 317–325
3. Bresin, R., Friberg, A.: Emotional Coloring of Computer-Controlled Music Performances. Computer Music Journal **24**(4) (2000) 44–63
4. Friberg, A., Bresin, R., Sundberg, J.: Overview of the KTH rule system for musical performance. Advances in Cognitive Psychology **2**(2) (2009) 145–161
5. Goebl, W., Dixon, S., Poli, G.D., Friberg, A., Bresin, R., Widmer, G.: ' Sense ' in Expressive Music Performance : Data Acquisition , Computational Studies , and Models. Artificial Intelligence (2005) 1–36
6. Kirke, Alexis, Miranda, E.R.: An Overview of Computer Systems for Expressive Music Performance. In: Guide to Computing for Expressive Music Performance. (2013) 1–47
7. Giraldo, S.I., Ramirez, R.: A machine learning approach to discover rules for expressive performance actions in jazz guitar music. Frontiers in Psychology **7** (2016) 1965
8. Bantula, H., Giraldo, S., Ramírez, R.: Jazz Ensemble Expressive Performance Modeling. Proc. 17th International Society for Music Information Retrieval Conference (2016) 674–680
9. Angulo, I., Giraldo, S., Ramirez, R.: Hexaphonic guitar transcription and visualization. In: TENOR 2016, International Conference on Technologies for Music Notation and Representation. (2016) 187 – 192
10. Cheveigne, A.D., Kawahara, H.: YIN, a fundamental frequency estimator for speech and music. **111**(April) (2002)