Modeling Embellishment, Timing and Energy Expressive Transformations in Jazz Guitar

Sergio Giraldo SERGIO.GIRALDO@UPF.EDU
Rafael Ramirez RAFAEL.RAMIREZ@UPF.EDU
Universitat Pompeu Fabra, Roc Boronat 138, 08018 Barcelona, Spain

Abstract

Professional musicians manipulate sound properties such as timing, energy, pitch and timbre in order to add expression to their performances. However, there is little quantitative information about how and in which context this manipulation occurs. This is particularly true in Jazz music where learning to play expressively is mostly acquired intuitively. In this paper we describe a machine learning approach to investigate expressive music performance in Jazz guitar music. We extract symbolic features from audio performances and apply machine learning techniques to induce expressive computational models for embellishment, timing, and energy transformations.

1. Introduction

Inarguably, music performance plays an important role in our culture. People clearly distinguish the manipulation of sound properties by different performers and create preferences based on these differences. However, there is little quantitative information about how and in which contexts expressive performance occurs. This is particularly true in Jazz music where most of the performance information is acquired intuitively.

Most of the past research in music expressive performance modeling has focused on classical music. Exceptions include the work by Lopez de Mántaras et al. and Ramirez et al. . Lopez de Mántaras et al. (Lopez de Mántaras, 2002) describe a system able to infer Jazz saxophone expressive performances from non-expressive monophonic descriptions using Case Based Reasoning. Ramirez et al. (Ramirez, 2006) applies inductive logic programming to obtain models capable of generating and explaining expressive Jazz saxophone performances.

5th International Workshop on Machine Learning and Music, Edinburgh, Scotland, UK, 2012.

In this paper, we investigate the manipulation of timing, energy, and ornamentation in Jazz guitar music in an attempt to understand and recreate expression in such performances. We describe the data collection process, the expressive transformations we model and report on preliminary results.

2. Jazz Guitar Expression Modeling

The training data used in this study are monophonic guitar recordings of standard Jazz pieces performed by a professional musician. The pieces were obtained from a jazz guitar book's accompanying CD (Marshall, 2000) in which the pieces' melodies were recorded on separate channels from the corresponding accompaniment backing tracks. The CD performances closely imitate original recordings of several jazz tunes by famous jazz guitar players. A complete transcription of each performance is included in the volume. We also obtained the inexpressive scores of the tunes from known jazz standard books compilations.

Note segmentation is performed using frequency and energy frame descriptors. The process relies on the implementation of the Essentia[1] audio processing library. Manual correction is performed after the onset detection process. The algorithm proposed by Dan Ellis (Dan Ellis, 2007) is used for beat tracking, and quantization of beat onsets is performed by linear regression. Chord information is parsed from Band in a Box files converted into text format. Both inexpressive and performance scores are transcribed to MIDI format and parsed using MIDI toolbox (Eerola 2004).

Each note in the training data is annotated with a number of attributes representing both properties of the note itself and some aspects of the context in which the note appears. Information about the note includes note

[1] Essentia & Gaia: audio analysis and music matching C++ libraries developed by the MTG (Resp.: N. Wack), http://mtg.upf.edu/technologies/essentia

duration, energy, pitch, and metrical position within a bar, while information about its melodic context includes information on neighboring notes (i.e. relative pitches and durations), as well as, melodic analysis with respect to the key and harmonic analysis with respect to the ongoing chord.

In Jazz, performers usually modify the original melody in order to embellish the melody. For each note of the score we manually searched its corresponding performance note(s). Thus, embellishments are characterized by the changes in duration, onset, and pitch with respect to the original note. This way we classify each note of the original score as *embellished* or *not embellished*, and generate a database of performed transformations.

We applied several machine learning algorithms (i.e. k-NN, model trees, artificial neural networks, support vector machines and decision trees) to induce duration, energy and embellishment transformation models. Once the embellishment model predicts a note to be embellished, we use k-NN to search for the closest embellishment note transformation in our database. The descriptor set for this search contains note duration, previous note duration, and metrical position within a bar.

Figure 1 shows the MIDI roll of the original inexpressive score (top), the performed score (center), and the predicted by the systems (bottom). Circles show the embellished notes, the performed and predicted embellishment.

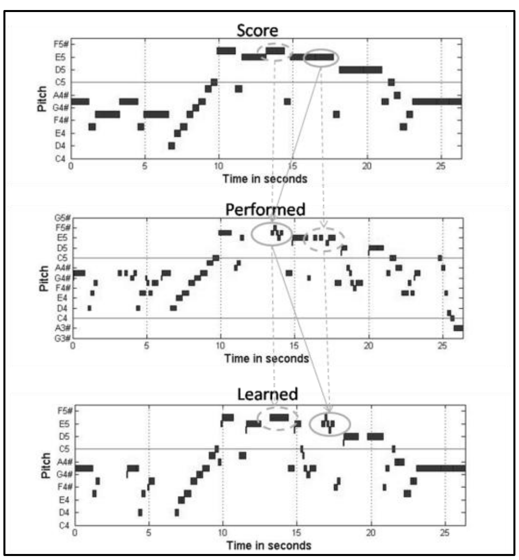

Figure 1. Piano roll of a melody fragment: Score (top), performed (center), and predicted (bottom)

3. Results

Lazy methods (k-NN and K*) were found to be the most consistent. Table 1 shows the correlation coefficients (CC) and the correctly classified instances percentage (CCI%) obtained with the different methods

Prediction Model	Duration Ratio (CC)	Energy Ratio (CC)	Embellishment (CCI%)
K*	0.4904	**0.7094**	**81,25**
k-NN (k=1)	**0.7431**	0.392	78,13
Model trees	0.2893	0.7226	-
ANN	0.5636	0.433	75,00
SVM	0.5915	0.3605	65,13
Decision trees	-	-	75,00

Table 1. Accuracy measures obtained.

Figure 2 shows the energy and duration ratio predictions obtained with k-NN and K*, respectively. The curves show duration deviation with respect to the score, and energy deviation with respect to the mean loudness.

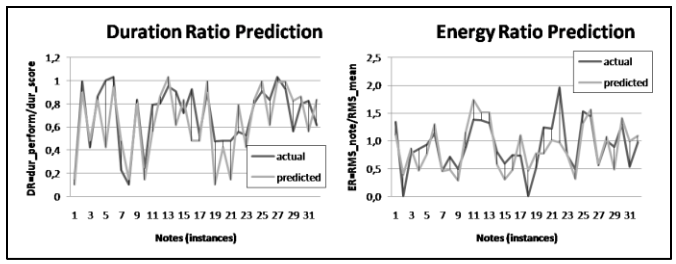

Figure 2. Duration and energy ratio predictions

4. Conclusion

In this study we have applied machine learning techniques in an attempt to recreate musical expression in jazz guitar music, by training models for duration (timing), energy and embellishment transformations. The preliminary results seem to indicate that the features extracted contain sufficient information allowing learning models to accurately capture the considered transformations.

References

Eerola, T. & Toiviainen, P. MIDI Toolbox: MATLAB Tools for Music Research. *University of Jyväskylä: Kopijyvä, Jyväskylä, Finland*, 2004.

Ellis, D. Beat Tracking by Dynamic Programming. *J. New Music Research*, Special Issue on Beat and Tempo Extraction, vol. 36 no. 1, March 2007, pp. 51-60. (10pp), 2007.

Lopez de Mantaras, R., Arcos, J.L. AI and music, from composition to expressive performance. *AI Mag*. 23 (3), 2002

Marshall, W. Best of Jazz Guitar. *Hall Leonard*, 2000.

Ramirez, R., Hazan, A.. A Tool for Generating and Explaining Expressive Music Performances of Monophonic Jazz Melodies, *International Journal on Artificial Intelligence Tools*, 15(4), pp. 673-691, 2006.