Towards automatic transcription of expressive oral percussive performances

Amaury Hazan

Music Technology Group, Pompeu Fabra University
Ocata 1, 08003 Barcelona, Spain, Tel:+34 935422165, Fax:+34 935422202
ahazan@iua.upf.es

1 Research Proposal

The research proposal presented here focuses on information retrieval in voice generated rhythms. The aim of this work is to develop a system able to reduce the gap between the user and a device (namely keyboard, drum pad or GUI) in order to get a symbolic rhythmic representation. This is relevant as many musicians who just have an intuitive notion of rhythm and groove cannot easily transcript a beat they have in mind. Furthermore, in both non western civilizations music and recent western urban genres, the oral tradition of music and especially rhythm is predominant. Few works have focussed in indexing automatically non standard drum based rhythms and an effort in finding a representation that can apply to a whole range of acoustic oral rhythms from beat boxing to Indian tabla oral recitals has to be done. This can be achieved in several ways, from finding an appropriate instrument taxonomy asserting that any performer tries to imitate some drum percussion, to using a completely data-driven model. Between theses extremes, one can consider the taxonomy of human phonemes as a starting point to identify recurrent oral drum clusters. The other issue raised by this work is that even if each percussive event was perfectly segregated and transcribed, the resulting drum score would lack of information describing how the performer has modulated the produced sounds. Thus, a representation of oral percussion expressive effects (e.g energy and resonance frequency variation, use of harmonic sounds, distortion) has to be defined along with effective computational methods to track them. We must provide enough flexibility to the system so that it is able to assign the same instrument to a family of somehow modulated sounds and at the same time being able to track the expressive modulation events. Recent works in non standard drum percussive signals transcription and sax expressive performance analysis lead to believe that we can take advantage of performing some contextual analysis of the oral drum part instead of just considering the description of each isolated percussive event.

2 Related work

Automatic audio analysis has been widely studied during the past years, and a growing interest has been shown in developing high level content extraction systems able to process automatically large amounts of data. Some works has largely helped the developments of the automatic drum classification branch. In [1] a system that performs sound segmentation using psychoacoustic knowledge (i.e using the characteristics of human auditive perception) has been designed. Automatic drum transcription has been studied in [2, 3] using the standard drum taxonomy at 3 levels of abstraction. These works were the main basis of [4],in which a system for oral percussion audio recognition was designed and implemented within the CLAM framework [1]. The system performed segmentation of input audio and then generates both temporal and spectral descriptors of the attack and decay part of each percussive fragment. Thus, each training vector was describing individual percussive events without referring to a larger context. A taxonomy of 4 standard drum sounds has been defined i.e the classes were Bass Drum, Snare Drum, Closed Hat and Open Cymbal. A training set formed with the recordings of 4 performers was used (they were asked to imitate the sound of each of the drum classes), totalizing more than 400 training instances, that is, more than 100 instances per drum class. A classification accuracy of 79% has been obtained in a test involving 2 new performers, the test set being composed of 150 instances. This work did not consider the problem of beat induction and meter detection, which was developed in [5, 6]. The problem of analyzing audio percussive excerpts which are not generated using standard drum sounds is studied in [8, 7] when some context analysis of the percussive events is considered. In [8], which focusses on the automatic labelling of tabla signals into tabla recital vocabulary, a Hidden Markov Model is considered to represent the contextual dependencies between percussive strokes and because the representation symbols may be context dependent. In [7], arbitrary sounds as voice, or taps can be segregated into 3 labels that represent a rhythmic role rather than an instrument, and thus concentrate on the pattern context of each rhythmical event. Melody context analysis has also been used in [9, 11, 12] to study expressive transformations in monophonic sax performances of jazz standards within ProMusic research project. Two approaches were considered, the first one in [9] used Case Base Reasoning. Given a test inexpressive melody, the system retrieves in a Knowledge Base the most similar one using a melodic similarity measure based on Narmour I/R structures ([10]), so that it is able to determine the expressive transformations which have to be performed. On the other hand, in our recent works, an Inductive Logic Programming approach is being considered i.e. the system tries to induce first order logic rules that take into account the melodic context of the notes to induce more o less general transformations rules, that can either be used to produce understandable but imprecise results [11], or generating new expressive melodies [12].

References

1. Klapuri, A. (1999). Sound Onset Detection by Applying Psychoacoustic Knowledge, Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP.

[1] http://www.iua.upf.es/mtg/clam

2. Herrera, P. Yeterian, A. Gouyon, F. (2002). Automatic classification of drum sounds: a comparison of feature selection methods and classification techniques. Proceedings of Second International Conference on Music and Artificial Intelligence Edinburgh, Scotland
3. Herrera, P. and Dehamel, A. and Gouyon, F. (2003). Automatic labeling of unpitched percussion Sounds, Proceedings of Audio Engineering Society, 114th Convention, Amsterdam, The Netherlands.
4. Hazan, A. (2004) Interfaz oral para el reconocimiento de ritmos. Master thesis. Facultat d'Informatica de Barcelona.
5. Gouyon, F. Herrera, P. (2003). A beat induction method for musical audio signals. Proceedings of 4th WIAMIS-Special session on Audio Segmentation and Digital Music London, UK.
6. Gouyon, F. Herrera, P. (2003). Determination of the Meter of musical audio signals: Seeking recurrences in beat segment descriptors. Proceedings of Audio Engineering Society, 114th Convention Amsterdam, The Netherlands
7. Jouni, P., Klapuri, A. (2003) Model-Based Event Labeling in the Transcription Of Percussive Audio Signals Proc. of the 6th Int. Conference on Digital Audio Effects (DAFX-03), London, UK, September 8-11.
8. Gillet, O. Richard,G. (2003). Automatic Labelling of Tabla Signals, Proc of ISMIR 2003, Baltimore, USA Oct. 2003.
9. Gómez, E. Grachten, M. Amatriain, X. Arcos, J. (2003). Melodic characterization of monophonic recordings for expressive tempo transformations. Proceedings of Stockholm Music Acoustics Conference 2003 Stockholm, Sweden
10. Narmour, E. (1990). The Analysis and Cognition of Basic Melodic Structures: The Implication Realization Model. University of Chicago Press.
11. Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2004). Understanding Expressive Transformations in Monophonic Musical Phrases. Submitted to SMC'04.
12. Ramirez, R. Hazan, A. (2004). Performing Expressive Transformations in Jazz Standards. Submitted to ICDM'04.

This article was processed using the LaTeX macro package with LLNCS style

Understanding players interaction in expressive transformations of jazz standards

Amaury Hazan

Music Technology Group, Pompeu Fabra University
Ocata 1, 08003 Barcelona, Spain, Tel:+34 935422165, Fax:+34 935422202
ahazan@iua.upf.es

Research Proposal Jazz is widely considered as being one of the most improvisational forms of Western Music. Within a jazz band the performers often interpret the original score more freely than classical musicians Alteration such as insertions, ornamentations, and deletions are very often performed in jazz. In [1, 2], we carry out a work aiming to understand and generate expressive transformations of monophonic sax performances of jazz standards. We obtained models of timing transformation (duration and onset deviations) and energy variation of the notes using Machine Learning techniques that considered first the local melodic context of each note of the score i.e. its relation in terms of duration and pitch with its neighbors along with their metrical strength within a bar. We are aware that the expressive transformations mentioned above are dependent of a more global context which describes both rhythmical and melodic structure of a piece and it was included a broader context representation of the sax melodies based on Narmour I/R structures in the background knowledge of the analysis system. We are also interested in considering in addition of the sax part context a representation of the score structure of the other instruments of the jazz band (namely drums, bass, guitar or keyboard for the chords). This would be a starting point to understand how the other instruments interact with the sax. A step forward is to build an expressive transformation model for these instruments each one considering in his background knowledge the score information of the others. One can consider that an interesting starting point is to build an expressive transformation model of the drums. Despite drum performance expressivity can not deal with note duration or pitch alterations, we believe that it is more related with onset deviations, note ornamentation, insertion and deletion events. Studying drum expressivity can help us to understand and model them, and requires the developpment rhythm-focussed representations. As mentioned in [3] improvisation has to be seen as a communication form, and the ultimate aim of this work is to study how the expressive transformation of an instrument performance can interact with the other instruments expressive transformation.

References

1. Ramirez, R. Hazan, A. Gómez, E. Maestre, E. (2004). Understanding Expressive Transformations in Monophonic Musical Phrases. Submitted to SMC'04.
2. Ramirez, R. Hazan, A. (2004). Performing Expressive Transformations in Jazz Standards. Submitted to ICDM'04.

3. Berliner, P.F. (1994) Thinking in Jazz : The Infinite Art of Improvisation. Chicago Studies in Ethnomusicology Series (CSE).

This article was processed using the LaTeX macro package with LLNCS style