Timbre-based Drum Pattern Classification using Hidden Markov Models

Michael Blaß

Institute of Musicology, University of Hamburg, Hamburg, Germany
m.blass@email.de.

Abstract. In order to explore the possibility of a timbre-based rhythm theory, a drum pattern classification system was developed, which is capable of describing the internal structure of a drum groove in a stochastic way. Using an onset detection algorithm, timbral features were extracted at every drum onset of the sample file. Next, a Hidden Markov Model (HMM) was fitted to the data. Local decoding of the model showed that the rate of correct classifications lies at 100 % when examining plain samples and decreases with advancing musical complexity. Furthermore, similar sounds were decoded differently.

1 Introduction

Todays common theories concerning musical rhythm fall short of taking its most prominent feature into account. They fail to describe rhythm in terms of sound. Modern popular music is built upon the fundamental concept of groove, which is established by different instruments, each having a specific timbre, playing together, but not necessarily at the same beats (Bader & Markuse, 1994). Assuming this, groove can be viewed as succession of distinct timbres, each affecting the following. Given rhythm works like this, a rhythm theory for analytical and composing tasks is needed, which is able to describe groove as a progression of timbral events. Since drum sets contribute a large part to groove in popular music, the first step towards a timbre-based rhythm theory was the development of a model capable of handling the sounds of a drum set before going into detail with multi-instrumental composition.

2 Onset Detector

The first thing to do was to develop an onset detection algorithm which is able to return the position of note onsets within a musical signal. In this approach, the normalized input signal is subdivided into N sections S of length l. Each section is then transformed to a pseudo-phase-space by setting up a two-dimensional mesh grid. The number of boxes the grid has is determined by the parameter m. The pseudo-phase-space of section S_i is then the set of coordinates $\mathcal{P}(\mathrm{S}_i) = \{(x_0, x_{0+d}), \ldots, (x_{k-d}, x_k)\}, x_k \in \mathrm{S}_i$ being the kth frame of S_i and $d \in \mathbb{N}$ specifying a delay of the index. Each coordinate lies in exactly one of the grid's

In MML 2013: International Workshop on Machine Learning and Music, ECML/PKDD, Prague, 2013.

boxes. The sum of points in each box is counted and divided by the total number of points, so that to each box a probability p_j is assigned. With these probabilities the information entropy can be calculated as $H(\mathrm{S}_i) = -\frac{1}{\log m}\sum_{j=0}^{m} p_j \log p_j$. After iterating over each section, the resulting list of entropy values mirrors the informational progression of the sample. Because of the chaotic behavior of transients, an onset shows up as a positive shift of entropy. Therefore it is assumed, that the local maxima in the list of entropy values are musical onsets. However, the parameters l, m and d have to be reset appropriately for every single sample file.

3 Timbral Features

Perception of timbre is essentially multi-dimensional (Grey & Moorer, 1977). Therefore, it can be viewed as a feature space comprising at least a spatial and a temporal dimension (Bader, 2013). In this approach, timbre was reduced to the perceptual dimension of brightness so as to handle it one-dimensionally. Brightness strongly correlates with the spectral centroid, which then was extracted at each onset position.

4 Trainig data

The sample data used are solo drum sections of Pop, Rock and Electro pieces. The pieces were fed to the system as 16 Bit PCM wave audio files, mono and of 44100 Hz sampling rate. The length of each sample ranges between 10 and 20 seconds. Additionally, one sample was synthesized in order to have a simple drum pattern with only one sound at the same time. This sample was generated using the Dance Kit of the Drum Kit selection of Apple's Garage Band.

5 The Hiddden Markov Model

Hidden Markov Models have widely been used in the scope of computational musicology (Temperley, 2007). The model proposed here comprises an m-state Markov Chain as unobserved state-dependent process and a Poisson Mixture Model as state-dependent process (Zucchini, 2009). For every sample, a Hidden Markov Model was trained on the corresponding spectral centroid data. The centroid values are assumed to be produced by a Poisson distribution of mean λ_i, each specifying a distinct timbral event. Note that, although the HMM's states are interpreted as instruments of the drum set in this paper, it is generally not assumed that each state corresponds to a distinct instrument. The transition probability matrix (t.p.m.) and the vector of means λ were estimated by direct maximization of the discrete log-likelihood, using the `R` programming language. Initial values for λ are chosen to evenly space the range between the minimal and maximal spectral centroid. The t.pm. is initialized with relatively large values on the main diagonal and values close to zero at the remaining positions. The result is a stochastic model of the samples rhythm described by means of its timbral progression. Local decoding of the model given the input data can be used to evaluate the model's goodness of fit. Furthermore, it is possible to sample a new time series of states from the model. Using an midi interface, a new grooves were generated, which resembles the original.

6 Results

The onset detecting performance goes up to 100 % when computing drum-only samples and diminishes with increasing musical complexity. Expectedly, most sounds were decoded according to the auditory impression, e.g all base drum sounds were subsumed under the same state, whereas all hi-hat beats were decoded to another. Furthermore, some surprising results could be observed. When applying an HMM to a sample that has less sounds involved than the HMM has states, it appears that two states are assigned to the same sound. Figure 1 displays the waveform of the synthesized sample, with the vertical bars marking the detected positions of onsets in the time series. To each of the HMM's states a color is assigned, with purple and yellow representing the hi-hat. One can see that hi-hat beats are consistently decoded to two different states regarding their metrical position. A reasonable interpretation could then define state purple as "hi-hat on 'two and' " and state yellow as "End of figure hi-hat". Furthermore, silence is also recognized as a timbral event, represented by state 1.

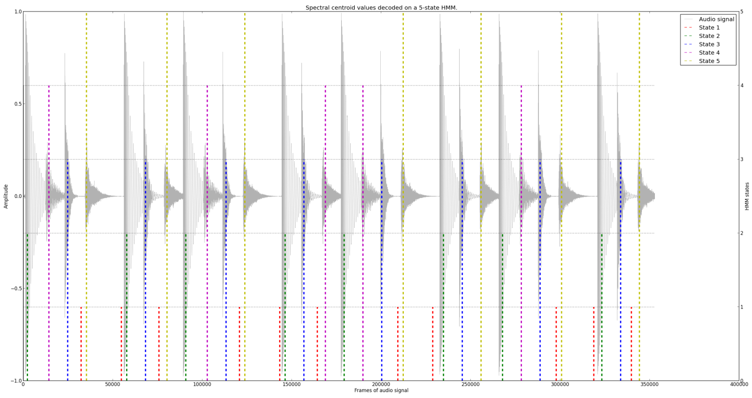

Fig. 1. The local decoding of this 4-state HMM shows how two similar sounds are distinguished regarding their position with the audio sample. State 1: Silence, State 2: Bass Drum, State 3: Snare, State 4: hi-hat on 2+, State 5: 'End of figure' hi-hat.

7 Conclusion

The model organizes similar sounds to different states, at the first glances, by means of their metrical position. Since the model does not learn any time related data, there have several hi-hat sounds to be involved. Although the sample was produced with only one hi-hat sound, a closer look at the wave form suggests that there is a difference in sound between state 4 and state 5 hi-hats. However, the difference was not audible. This could be explained by a slight overlap of the bass drum and hi-hat sounds, adding low frequencies to the spectral centroid measurement at each hi-hat onset following a bass drum. This indicates that the model is capable of detecting even subtle distinctions of sound. Nevertheless, both hi-hat states are not distributed randomly in the sample. Each is obtained on a designated metrical position. Therefore, a distinction in terms of position happens even though only timbral features were computed. Enhancing this model, a system is possible showing how instruments of a drums are played differently regarding their position within a bar. Furthermore, this could lead to a detection method of playing styles.

References

Bader, R., Markuse, B.: Perception and Analyzing Methods of Groove in Popular Music. In *Proceedings of the International Conference for Music Perception and cognition.* (1982) 401–402

Bader, R.: Nonlinearities and Synchronization in Musical Acoustics and Music Psychology. (2013)

Grey, J.M., Moorer, J.A.: Perceptual evaluations of synthesized musical instrument tones. Journal of the Acoustical Society of America **62**, (1977) 454–462.

Temperley, D.: Music and Probability. (2007)

Zucchini, W., MacDonald, I. L.: Hidden Markov Models for Time Series. (2009)