Stochastic Modeling of Expressiveness: Representing the Temporal Evolution of the Descriptors Using HMM

Gabriel Pereira Pezzini
Center for the Study of Musical Gesture and Expressiveness (CEGeME), Federal University of the Minas Gerais state (UFMG), Brazil. http://cegeme.musica.ufmg.br/

Thiago A. M. Campolina
Center for the Study of Musical Gesture and Expressiveness (CEGeME), Federal University of the Minas Gerais state (UFMG), Brazil. http://cegeme.musica.ufmg.br

Maurício Álves Loureiro
Center for the Study of Musical Gesture and Expressiveness (CEGeME), Federal University of the Minas Gerais state (UFMG), Brazil. http://cegeme.musica.ufmg.br

In: Jakubowski, K., Farrugia, N., Floridou, G.A., & Gagen, J. (Eds.)
Proceedings of the 7th International Conference of Students of Systematic Musicology (SysMus14)
London, UK, 18-20 September 2014, http://www.musicmindbrain.com/#!sysmus-2014/cfmp

Different interpreters do not play identically during a music performance, introducing their own expressive features. Although these features are perceptually recurrent for each musician, the deterministic modeling is a difficult task, making it more interesting to model by a stochastic patterns approach. This paper aims to model the temporal evolution of the acoustic features using HMM (Hidden Markov Model) note-by note, intrinsically related to the expressive intent of the artist performing the musical fragments. Descriptors related to changes in dynamics, tempo, attack and release have been implemented and tested. Dynamics was described by the RMS (Root Mean Square) energy changes for each note in comparison with the previous. Tempo was described by the IOI (Inter Onset Interval) deviation normalized by the note duration according to the score. Finally, the attack and release were described by the logarithm of their duration time. The methodology is divided into two parts: first the optimal number of states for each HMM is determined by taking the maximum of the curve between number of states and recognition rate. In the second, the recognition achieved by optimal number of hidden states is analyzed together with the structure of the probabilities of transitions between the states matrix. The training and testing data used in recognition tests were executions of the same fragment for clarinet of the fourth movement of Mozart's Quintet. Five musicians performing six times each were recorded. The results indicated better recognition rates using tempo, dynamics and attack time descriptors, reaching 60 hitting percent, with two, six and five hidden states respectively.

Introduction

Interpreters do not play identically during a music performance, introducing than their own expressive features. The comprehension of the processes involved in the production and perception of an expressive performance has been approached by models that quantify the player's expressive intentions, based upon musical expressiveness have demonstrated that musicians use small variations of duration, articulation, intensity, pitch and timbre to communicate to the listener aspects of the music that they interpret (Gabrielsson, 2003; Juslin, 2000). By comparing performances of different musicians as well as different interpretations by the same musician, these deviations can be perceived with a surprising clarity, even by non-specialised listeners. The quantification of the interpreter's expressive intentions upon such deviations involves the identification and measurement of a set of descriptor parameters defined and calculated from information extracted from the audio signal of the recorded musical performance. In order to incorporate different aspects of the resources used by performers to communicate their intention of expressiveness and intelligibility, these parameters may be established for different segmentation levels, as described below.

The Aims

This study aims to represent the temporal evolution of the acoustic features, note-by note, intrinsically related to the expressive intent of the artist performing the musical excerpts, through stochastic modeling using HMM (Hidden Markov Model). A recognition test was developed as a reliability measurement of the models.

Data Preparation

The first step for an appropriate estimate of descriptor parameters is to segment the signal into events to be analyzed, such as musical notes, note groups, or specific regions within a single note.

Segmentation

Segmentation is not a trivial problem, even on monophonic musical signals, especially if we consider the subjectivity in the discrimination of these events. Note onsets and offsets were detected on the RMS envelope averaged for 23 ms using an adaptive threshold, as suggested by De Poli (1998), calculated as the average energy in a certain neighborhood (1 s for a step of 6 ms) of each point of the RMS. Onset and offsets are detected by searching the minimum RMS between two consecutive values crossed by this dynamic threshold. The estimate of the fundamental frequency changes, with a pitch threshold below a half tone, helped the segmentation in cases where the detection of onsets and offsets was not possible by means of energy level only, such as legato notes. The end of attack was defined as the first amplitude maximum after the note onset, and the beginning of release, as the first amplitude maximum before the note offset. These points were detected by searching for maximum variations of the first derivative of the RMS signal. This definition of attack is adequate to describe the attack in most situations, but further consideration was necessary in cases where maximum amplitude was reached much later in the sustained segment of the note. The presence of transients during note transitions makes possible the use of spectral flux to detect the end of attack, since this point can be related to the reestablishment of harmonics amplitude correlation after note transitions (Loureiro, 2009).

Descriptors

Four descriptors related to changes in dynamics, tempo, attack and release have been implemented. Dynamics were described by the RMS (Root Mean Square) energy proportion changes for each note in comparison with the previous. Tempo was described by the IOI (Inter Onset Interval) deviation normalized by the quarter note duration according to the score. Finally, the attack and release were described by the logarithm of their duration time.

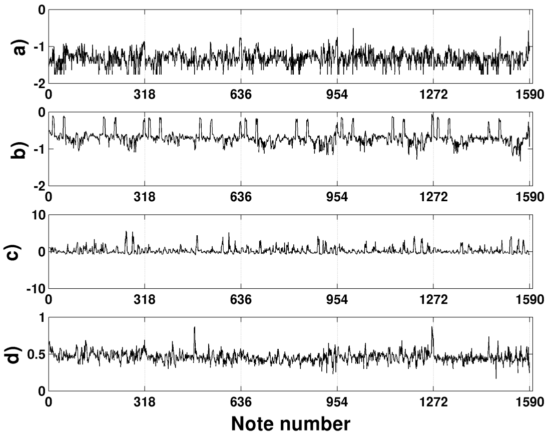

Figure 1. Data set shown in sequence for The vertical dotted lines mark musician changes. a) logarithm of attack time, b) logarithm of release time, c) Dynamic proportion and d) IOI normalized time, by the quarter note (0.5 seconds).

Quantization

In order to enable the same model structure to all the descriptors and musicians, the data set was normalized and quantized in 20 values. It was accomplished by subtracting the mean, dividing by the range and quantized in 20 values. Figure 2 shows the result of quantization.

Modeling

The modeling strategies was defined according to Rabiner and Juang (1986). It was assumed there was one HMM to each musician and descriptor. Each recording was assumed as one model emission, and the hidden states as internal mechanisms used to produce the expressiveness changes.

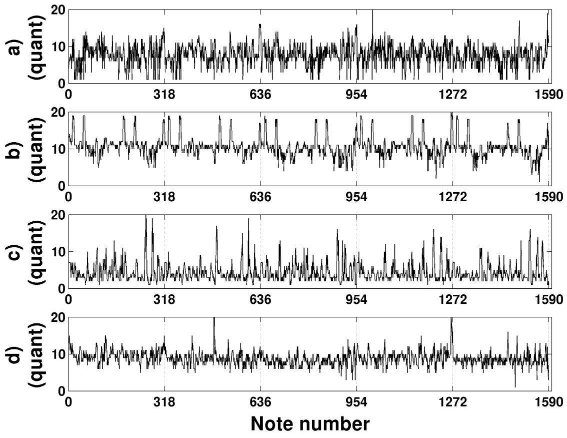

Figure 2. Quantized data set shown in sequence for The vertical dotted lines mark musician changes. a) logarithm of attack time, b) logarithm of release time, c) Dynamic proportion and d) IOI normalized time.

Training and Testing

The Baum-Welch algorithm was used to train the HMM's, while the Forward-Backward algorithm was utilized to decode the emissions. After the training of one model for each combination descriptor-musician, we have 20 models. For each, descriptor we have five test recordings and five HMM's to test the recognition. At this time, the decode algorithm is applied to each combination HMM-emission to return the log probability of the sequence be a emission of the model. The highest log probability then is chosen to indicate the generator HMM, and thus confront the right origin.

Results

In order to measure the reliability of the models trained, we proposed a recognition test. The test was investigated along the growth of the HMM hidden states number.

Data Description

The database is formed by five clarinetists performing six times, where five each were used to train the model, while one each was used to test. So we have a total with 30 recordings, with 53 notes each. The excerpt used was a fragment for clarinet of the fourth movement of Mozart's Quintet, shown in Figure 3 below.

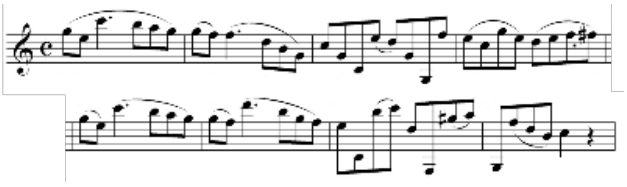

Figure 3. Fragment of the fourth movement of Mozart's Quintet used in the tests.

Recognition Test

The recognition for IOI, Dynamics changes and the logarithm of attack time reach 60 per cent hit-rate, while the logarithm of release time, reaches a lower hit-rate of only 2 correct guesses out of the five tested models. The hitting along the growth of the HMM hidden states number can be seen in Figure 4, and Table 1 summarizes the result.

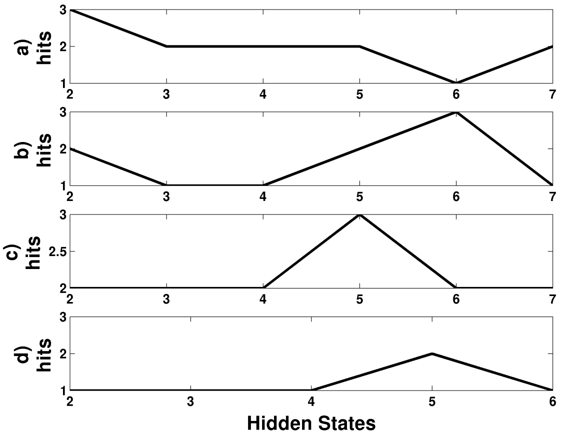

Figure 4. Recognition hits according to the model hidden states growth. A total of five excerpts was tested, one for each musician. In a) we have IOI, b) Dynamics proportion chances with the previous note, c) logarithm of attack time and d) logarithm of release time.

Descriptor	hits	Hidden states number
IOI	60%	2
Dynamics	60%	6
Attack	60%	5
Release	40%	5

Table 1. Number of hidden states that provides the highest hate of the recognition test. Five clarinetists performed six times, where five each were used to train the model, while one each was used to test.bss

Discussion

This study presents a quantitative way to represent time evolution of descriptors changes, intrinsically related to the expressive intent of the artist performing the musical fragments. It is not within the scope of this paper, but this information could be used to infer some possibilities for how musicians make decisions about changes in the descriptors, as IOI presenting two hidden states, could represent an internal mechanism for speeding-slowing tempo.

Acknowledgments. The accomplishment of this work was possible thanks to the participant clarinetists. This work was supported in part by FAPEMIG (Research Funding Agency of Minas Gerais State, Brazil), CNPq (Brazilian National Council for Scientific and Technological Development) and CAPES (Coordination for the Improvement of Higher Education Personnel).

References

De Poli, G.; Roda, A.; e Vidolin, A. (1998) *Note-by-Note Analysis of the Influence of Expressive Intentions and Musical Structure in Violin Performance.* Journal of New Music Research, v. 27, n. 3, p. 293-321.

Gabrielsson, A. (2003) *Music Performance Research at the Millenium*, Psychology of Music, vol. 31, pp. 221-272.

Juslin, P. N. (2000) *Cue utilization in communication of emotion in music performance: relating performance to perception.* Journal of Experimental Psychology: Human perception and performance, vol. 26, pp. 1797-1813.

Loureiro, M. A; Yehia, H. C.; Paula, H. B.; Campolina, T. A. M.; Mota, D. A. (2009) *Content analysis of note transitions in music performance.* Proceedings of the 6th Sound and Music Computing Conference, Porto, Portugal.

Rabiner, L.; Juang, B. H. (1986) *An introduction to hidden Markov models.* ASSP Magazine, IEEE, vol. 3, pp. 1-16.