RECOGNITION OF ISOLATED MUSICAL PATTERNS USING DISCRETE OBSERVATION HIDDEN MARKOV MODELS

Aggelos Pikrakis[1], *Sergios Theodoridis*[1], *Dimitris Kamarotos*[2],
[1] Department of Informatics, Division of Communications and Signal Processing
University of Athens, e-mail pikrakis@di.uoa.gr, stheodor@di.uoa.gr
[2] IPSA-Aristotle University of Thessaloniki, e-mail dimik@forum.ars.net.gr

ABSTRACT

Recognition of pre-defined musical patterns is very useful to researchers in Musicology and Ethnomusicology. This paper presents a novel efficient method for recognizing isolated musical patterns, using discrete observation Hidden Markov Models. The first stage of our method is to extract a vector of frequencies from the musical pattern to be recognized. This is achieved by means of a combination of a moving window technique with a largest-Fourier-peak selection algorithm. Each extracted peak frequency is subsequently quantized to a symbol of a finite and discrete alphabet. The resulting sequence of quantized frequencies is given as input to a set of Hidden Markov Models (HMM). Each HMM has been trained to model a specific pre-defined musical pattern. The unknown musical pattern is assigned to the model which generates the highest recognition probability. We have applied our method for the recognition of isolated musical patterns in the context of Greek Traditional Music. The resulting recognition rate was higher than 95%. Greek Traditional Music has distinct features, which distinguish it from the western equal-tempered tradition. *To our knowledge, this paper presents a first effort for musical pattern recognition in the context of Greek Traditional Music using discrete observation Hidden Markov Models. Previous work was based on Dynamic Time Warping [8] .*

1 INTRODUCTION

Digital recording techniques and specifically audio data storage, combined with digital sound processing tools, offer new possibilities in analysis of musical structures, the modeling of acoustic characteristics of an instrument and the musical pattern comparison and recognition.

An essential condition in this new approach is the existence of digital sound processing tools that will permit assisted semi-automated search of sound structures, within large number of stored sound files. These sound structures are rich in information related to important musicology tasks. Examples of such tasks include the spectral characteristics of an instrumental playing mode, the transitory musical patterns etc. The musical patterns have been shaped and categorised through practice in many musical traditions. Our proposed recognition scheme is based on the extraction of a vector of frequencies from the input musical pattern, using a moving window technique. For each window a largest-Fourier-peak algorithm is applied. The extracted vector of peak-frequencies is given as input to a recognizer, which employes discrete observation Hidden Markov Models in order to determine the input pattern.

2 EXTRACTION OF THE VECTOR OF FREQUENCIES

The first stage of our method is to extract a vector of frequencies from the unknown musical pattern. Specifically, a moving window Fourier Transform is performed on the musical pattern. For each window the frequencies corresponding to the two most dominant peaks are selected as feature candidates. From these two frequencies, the lowest is selected as the respective feature if a) it corresponds to the most dominant peak or b) it corresponds to a peak, whose amplitude is higher than a percentage threshold relative to the most dominant peak. The value of this threshold was set equal to 0.25 after extensive experimentation. In any other case the higher of the two frequencies is selected as the respective feature. Furthermore, if the highest Fourier peak of a frame is less than a pre-defined threshold, the frame is considered to be noisy and the extracted frequency is set equal to one. Figure 1 shows the evolution of the frequency content of a musical pattern over time. It demonstrates that the frequency content of the signal is split into horizontal frequency bands. Ideally, the dominant frequency should always be found on the same band. It would then suffice to choose the frequency corresponding to the dominant peak from each frame. However, this is not always the case. It turns out, that the dominant peak can be found in either of the two lowest bands. So, the algorithm should always track the lowest distinguishable frequency band, even if for some frames the dominant peak isn't located in that band.

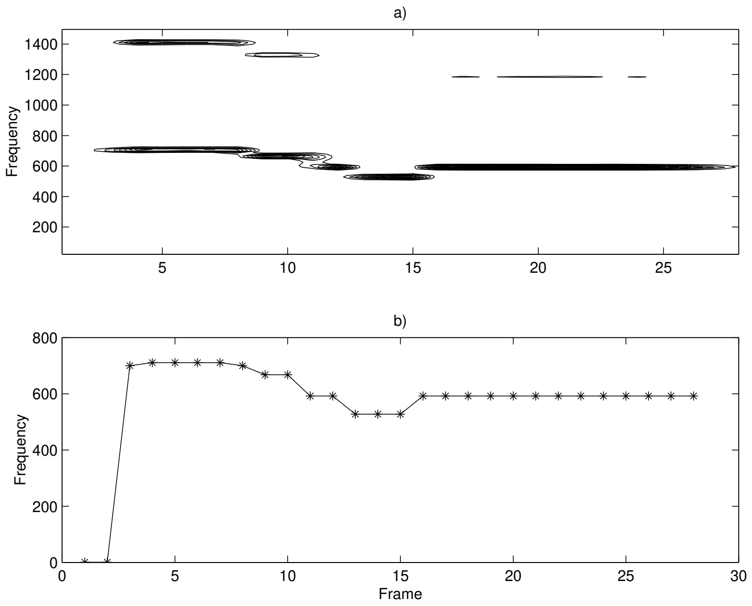

Figure 1: a) Contour plot of the spectrogram of a musical pattern b) The extracted frequency vector.

3 QUANTIZATION OF THE EXTRACTED FREQUENCIES

Let $f_i, i = 1 \ldots M$ be the previously extracted sequence of frequencies. As a post-processing step, the logarithm of f_i is calculated $\forall i$, i.e, $l_i = \log f_i, i = 1 \ldots M$. This is because we are trying to imitate some aspects of the human auditory system, which is known to analyse an input pattern using a logarithmic frequency axis. Each l_i is subsequently mapped to a symbol v_i of a discrete finite alphabet V using the following mapping function:

$$v_i = \lfloor \frac{l_i - S}{step} \rfloor \ , \ v_i \in V$$

where S and $step$ are constants. S corresponds to the lowest possible value that l_i may take and $step$ is equal to the width of the fixed size bins to which the range of possible values of l_i is divided. For the case of the musical patterns that we studied we used an alphabet of fifty discrete symbols. S and $step$ were set to 2.1 and 0.02 respectively. Figure 2a demonstrates a sequence of extracted frequencies before quantization has taken place and Figure 2b the same sequence after quantization has been applied.

4 DISCRETE OBSERVATION HMMs

The previously extracted sequence $v_i, i = 1 \ldots M$, of discrete symbols is given as input to a set of discrete observation HMMs. We use the notation $\lambda = (N, \ V, \ A, \ B, \ \pi)$ to represent a discrete observation HMM, where N is the set of states of the model, V is the discrete set of possible symbols, A is the state transition probability distribution, B is the observation symbol probability distribution for the set N and π is the initial state distribution. Each HMM has been trained to recognise a specific musical pattern. For each HMM, we compute the probability that it has generated the sequence of observations $v_i, i = 1 \ldots M$. The HMM corresponding to the largest probability is considered to be the model which best matches the observation sequence v. For the calculation of the above probabilities we experimented with both the forward-backward and the Viterbi approaches. The results were similar in both cases.

4.1 TRAINING OF THE HMMs

Each HMM was individually trained to recognise a specific type of musical pattern. For each type of musical pattern a set of patterns was chosen for the training phase. The model parameters $(A, \ B, \ \pi)$ were adjusted using two different approaches: the Baum-Welch and the Viterbi reestimation formulas. Similar training performance was observed for the two methods. However, in some occasions, the Baum-Welch approach converged more rapidly.

5 AN APPLICATION OF OUR METHOD IN THE CONTEXT OF GREEK TRADITIONAL MUSIC

The musical system of Greek Traditional music and the techniques of instrument players give to this sound material *a radicaly different structure when compared with this of western equal - tempered tradition.* Some major differences are :

- the use of larger, formalised transitory patterns as a main element of the musical structure and not as an ornamental one, as it is in the case of western musical tradition.

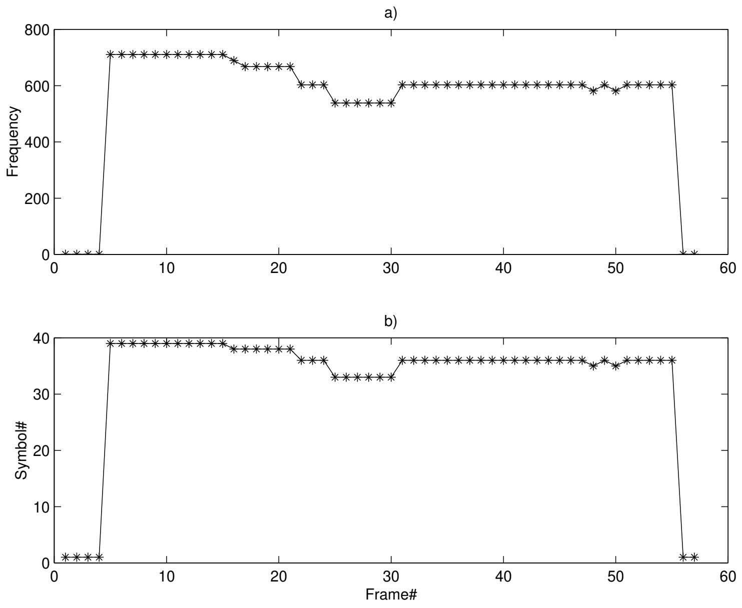

Figure 2: a) Sequence of extracted frequences prior to quantization b) Same sequence after quantization.

- the existence of transitional patterns between notes,
- the intervalic system (system of musical scales) that contains many musical intervals that are smaller than the well-tempered.

From a large number of transitory musical patterns in different instrumental styles, we have selected the ten most typical models of groups usually encountered in practice. The criteria for this choice were the common use of these patterns in various Greek traditions and the differences on time length that occur in their use. This last musical characteristic -the elasticity of the musical pattern, retaining its musical function, while stretching its total length up to five times in some cases- is a major problem when studying this material. A set of 500 musical patterns were generated by four Greek Traditional Clarinet players in a monophonic environment. A small subset of the above patterns was used to train 10 discrete observation HMMs as described in the previous section. We experimented with both left-to-right and ergodic HMMs and trained the models with both the Baum-Welch and Viterbi methods. All experiments were carried out using Matlab. In all cases the overall recognition rate was above 95%.

6 FUTURE RESEARCH

In the future we will focus our research activity on the application of continuous observation HMMs and Hybrid Models for the same recognition task. Hybrid Models consist of a combination of continuous observation HMMs with neural networks and have already been applied with success in the context of speech recognition.

References

[1] J. Deller, J. Proakis, J. Hansen, "Discrete-Time Processing of Speech Signals", McMillan, 1993.

[2] H.F. Silverman, D.P. Morgan, " The Application of Dynamic Programming to Connected Speech Recognition", IEEE ASSP Magazine, July 1990.

[3] J.C Brown, " Musical fundamental frequency tracking using a pattern recognition method", J. Acoust. Soc. Am. 92(3), September 1992.

[4] S.Karas "Theoritikon: Methodos on Greek Traditional Music", ed. Athens 1982, (in Greek).

[5] D.R. Stammen, B. Pennycook, "Real-time Recognition of Melodic Fragments using the Dynamic Time Warping Algorithm", Proceedings of the Int. Computer Music Conference, 1993.

[6] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", Proceedings of the IEEE, Vol. 77, No. 2, February 1989.

[7] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns in the context of Greek Traditional Music", Procedings of the IEEE ICECS, 1996.

[8] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns in the context of Greek Traditional Music using Dynamic Time Warping Techniques", Proceedings of the Int. Computer Music Conference, 1997.