

CLASSIFICATION OF MUSICAL PATTERNS USING VARIABLE DURATION HIDDEN MARKOV MODELS

Aggelos Pikrakis, Sergios Theodoridis and Dimitris Kamarotos

Department of Informatics and Telecommunications, University of Athens
Panepistimioupolis, TYPA Buildings, Zografou 15784, Athens, Greece
phone: +30 210 72725328, fax: +30 210 7275337, email: pikrakis@di.uoa.gr
web: www.di.uoa.gr/dsp

ABSTRACT

This paper presents a new extension to the variable duration Hidden Markov model, capable of classifying musical patterns that have been extracted from raw audio data, into a set of predefined classes. Each musical pattern is converted into a sequence of music intervals by means of a fundamental frequency tracking procedure and it is subsequently given as input to a set of variable duration Hidden Markov models. Each of these models has been trained to recognize patterns of the respective predefined class. Classification is determined based on the highest recognition probability. This new type of variable duration Hidden Markov model provides increased classification accuracy because a) it deals effectively with errors originating during the feature extraction stage and b) it accounts for variations due to the expressive performance of instrument players. To demonstrate its effectiveness, the novel classification scheme has been employed in the context of Greek traditional music, to monophonic musical patterns of a popular instrument, the Greek Traditional clarinet. The classification results demonstrate that the new approach outperforms previous work based on conventional Hidden Markov models.

1. INTRODUCTION

Algorithms for the effective comparison of musical patterns have gained increased interest in recent years due to their use in content-based music retrieval applications, including query-by-humming and repeating pattern finding systems. Most research efforts so far, have mainly concentrated on comparing MIDI signals [1], which can be a severe limitation for a large number of real world problems involving raw audio signals.

This paper provides a solution to the problem of matching an unknown monophonic musical pattern that has been extracted from raw audio data, against a set of predefined musical patterns, each of which is represented by a variable duration Hidden Markov model (HMM) [2]. The novelty of the approach lies in the fact that a new modified Viterbi algorithm [3], [4] is proposed for the calculation of the recognition probability generated by each variable duration HMM. This algorithm provides increased recognition performance compared to previous work [5], that was based on standard HMMs, while reducing the required number of states per HMM. The modified Viterbi algorithm provides a solution for the classification of instances of a musical pattern that deviate from a prototype pattern, due to errors arising in the feature extraction stage and due to the performance variations of the instrument players. These errors are very common in practice. A training algorithm for the HMMs is also introduced, in the light of the new modified Viterbi algorithm and a methodology for the construction of the HMMs is also presented. The use of variable duration HMMs permits to circumvent a major weakness of conventional HMMs, i.e., the modeling of state duration.

Section 2 describes the feature extraction stage, which extracts a sequence of music intervals from the raw audio data, by means of a

fundamental frequency tracking algorithm followed by a quantizer. Section 3 presents the modified Viterbi algorithm for the calculation of the recognition probabilities, along with the training algorithm and the methodology for building the HMMs. A case study involving the application of the proposed scheme in the context of Greek Traditional music is presented in section 5. The last section presents our conclusions and future research priorities.

2. FEATURE EXTRACTION

The goal of the feature extraction stage is to provide a time-frequency representation of the unknown musical pattern, to be used as input to a set of variable duration HMMs. At a first step, a sequence of fundamental frequencies is extracted from the unknown musical pattern. For this task, any robust fundamental frequency tracking algorithm can be employed, like [7], [8], [9]. Let $\mathbf{F} = \{f_1, f_2, \dots, f_M\}$ be the sequence of extracted fundamentals, where M is the number of frames that the signal is split into, by means of a moving window technique. During this step, a number of errors are likely to occur. Such is the case when a transition between notes takes place. We chose not to apply any heuristic rules to reduce the number of errors, but rather treat these errors directly by the variable duration HMMs.

At a second step, in order to imitate certain aspects of the human auditory system, which is known to analyze an audio pattern on a logarithmic frequency axis, each f_i is mapped to a positive number, say k , equal to the distance (measured in quarter-tone units) of f_i from f_0 (the lowest fundamental frequency of interest), i.e., $k = \text{round}(24 \log_2 \frac{f_i}{f_0})$, where $\text{round}(\cdot)$ denotes the roundoff operation. As a result, sequence \mathbf{F} is mapped to sequence $\mathbf{L} = \{l_i, i = 1 \dots M\}$, where l_i lies in the range 0 to some maximum value, say l_{max} .

It is now straightforward to compute \mathbf{D} , the sequence of music intervals (frequency jumps), from the sequence \mathbf{L} . This is achieved by calculating the difference of sequence \mathbf{L} , i.e., $\mathbf{D} = \{d_i = l_{i+1} - l_i, i = 1 \dots M - 1\}$. Most of the time l_{i+1} is equal to l_i because each note in a musical pattern is likely to span more than one consecutive frames. As a result, $d_i = 0$ for most of the frames (i 's). By calculating differences, we deal with the fact that instances of the same musical type may have different starting frequencies. We assume that the d_i 's fall in the range of $-G$ to G , where G is the maximum allowed music interval. This feature extraction scheme has also been used in [6].

3. THE VARIABLE DURATION HMM

Let us first rewrite sequence \mathbf{D} as

$$\mathbf{D} = \{\mathbf{0}_{z_1}, m_1, \mathbf{0}_{z_2}, m_2, \dots, \mathbf{0}_{z_{N-1}}, m_{N-1}, \mathbf{0}_{z_N}\}$$

where $\mathbf{0}_{z_k}$ stands for z_k successive zeros (i.e., zero valued d_i 's) and each m_i is a non-zero d_i . This alternative representation reveals the fact that sequence \mathbf{D} can be considered as consisting of subsequences of zeros separated by positive values (the m_i 's), with each m_i denoting a music interval, i.e. the beginning of a new note. The physical meaning of a subsequence of zeros is that it represents a

This work was partly supported by the EU project "PANDA" (IST-2001-33058)

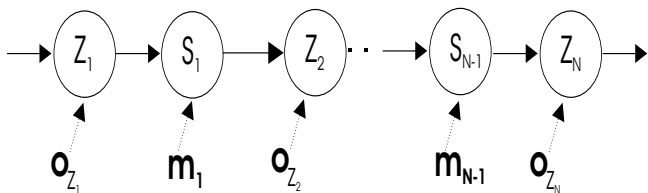


Figure 1: A discrete observation HMM that models instances of the same musical pattern, in the ideal case

steady musical note. The length of the subsequence, measured in frames is actually the duration of the note. In the ideal case, \mathbf{O}_{z_1} corresponds to the first note perceived by the human ear, m_1 is the music interval equal to the difference (measured in quarter tones) of the first two notes, \mathbf{O}_{z_2} corresponds to the second note and so on, until m_{N-1} is reached, corresponding to the last music interval, followed by \mathbf{O}_{z_N} , the duration of the N th (i.e. last) note perceived by the human ear.

3.1 The ideal case

In the ideal case, instances of the same musical pattern should only differ in the number of zeros separating the m_i 's. The HMM chosen to model each pattern, follows the above structure of the feature sequence. That is, each note expected to be in the pattern, corresponds to a number of states, namely a non-zero state followed by a zero-state, with the exception, of course, of the first note. Each subsequence \mathbf{O}_{z_k} can be emitted by a specific state of the HMM and each m_i is assigned to a state of each own (figure 1). As a result, for a pattern consisting of a sequence of N notes, the respective HMM consists of $S = N + N - 1 = 2N - 1$ states. For notational purposes, the HMM states corresponding to the zero valued subsequences are named Z-states, Z_1, \dots, Z_N and the states corresponding to m_i 's are the S-states, S_1, \dots, S_{N-1} (figure 1). The reason different zero-states are used, is that this allows a different state duration model to be adopted for each state, something that is dictated by the nature of our signals.

Translated in the HMM terminology, let $\mathcal{H} = \{\pi, A, B\}$ be the resulting HMM, where $A_{S \times S}$ is the state transition matrix, $B_{(2G+1) \times S}$ is the symbol probability matrix (G is the maximum allowed music interval), and $\pi_{S \times 1}$ is the vector of initial probabilities. For the moment we assume that Z_1 is always the first state, therefore $\pi(1) = 1$ and $\pi(i) = 0, i = 2 \dots S$.

For the ideal case, each Z-state can only emit zeros and each S-state can only emit the respective m_i . Therefore, there is only one positive element at each column of matrix B with a value equal to one.

So far, we have made no effort to model explicitly the time duration of each state, which is also the major weakness of conventional HMMs. In order to circumvent this weakness, we resort to the so-called variable duration HMM which is capable of modeling explicitly the time spent at each state, i.e., the expected number of symbols emitted by each state. We assume that for each Z-state, duration is modeled by a single gaussian probability distribution, $p_{Z_i}(\tau) = \mathcal{G}(\tau, \mu_{Z_i}, \sigma_{Z_i}^2)$, where $1 \leq \tau \leq T$, T being a constant. The values of T , μ_{Z_i} and σ_{Z_i} can be chosen to reflect the amount of permitted time warping. Since the S-states can only emit one symbol at a time, we can assume for the time being, that the respective gaussian distributions have a mean value equal to 1 and small σ values.

3.2 Deviations from the ideal case - The need for a modified Viterbi algorithm

In practice, N , the number of extracted notes, is rarely equal to the number of notes perceived by a human listener, mainly due to errors in the feature extraction stage. For example, let us consider the contour plot of the spectrogram of a musical pattern played by a wind instrument, the clarinet, as shown in figure 2. This pattern consists

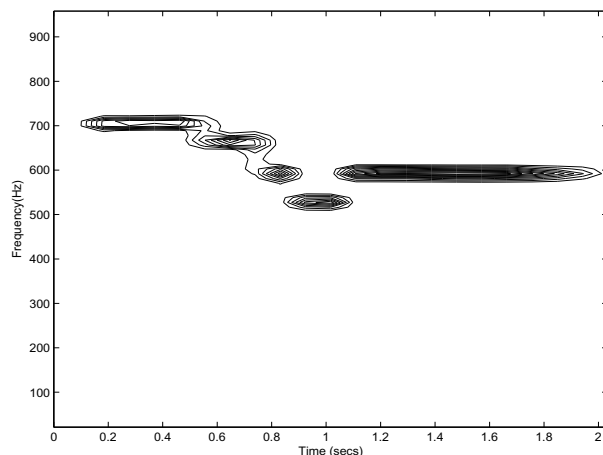


Figure 2: Contour plot of the spectrogram of a musical pattern played by Greek Traditional clarinet

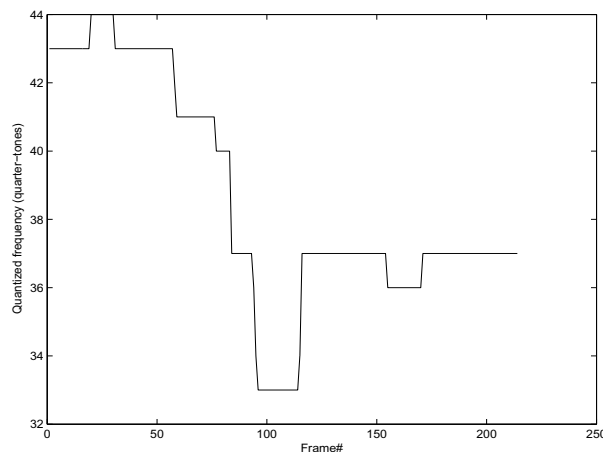


Figure 3: Quantized fundamental frequencies corresponding to the musical pattern of figure 2, having employed Brown's narrowed autocorrelation method.

of five notes and assuming a quarter-tone resolution is adopted (as is the case for the signals that we studied), the \mathbf{D} sequence for this pattern possesses the structure

$$\mathbf{D} = \{\mathbf{0}_{z_1}, -2, \mathbf{0}_{z_2}, -4, \dots, \mathbf{0}_{z_3}, -4, \mathbf{0}_{z_4}, 4, \mathbf{0}_{z_5}\}$$

where $\mathbf{0}_{z_k}$, $k=1 \dots 5$, depends on the length of the moving window and the overlap between successive frames. Figure 3 shows the fundamental frequency tracking results after quantization has taken place, having used Brown's narrowed autocorrelation fundamental frequency tracker. It can be seen that the extracted symbol sequence \mathbf{D}_t deviates from what is expected and has the structure $\mathbf{D}_t = \{\mathbf{0}_{z_1}, 1, \mathbf{0}_{z_2}, -1, \mathbf{0}_{z_3}, -2, \mathbf{0}_{z_4}, -3, \mathbf{0}_{z_5}, -1, \mathbf{0}_{z_6}, -4, \mathbf{0}_{z_7}, 4, \mathbf{0}_{z_8}, -1, \mathbf{0}_{z_9}, 1, \mathbf{0}_{z_{10}}\}$. If \mathbf{D}_t is given as input to an HMM that was built following the ideal case, a zero recognition probability would occur, which is clearly undesirable.

On the other hand, careful observation reveals that m_1 , which is equal to 1 and m_2 , which is equal to -1 , cancel out and so do m_8 and m_9 . In addition, $m_4 + m_5 = -4$ which is the music interval actually perceived by the human ear. Previous work by the authors [5] has made an effort to categorize such deviations from the ideal case. Three broad categories were discerned:

- (a) subsequences of the form $\{+1, \mathbf{0}_{z_k} - 1\}$ or $\{-1, \mathbf{0}_{z_k} + 1\}$ where $+1$ and -1 cancel out. In the more general case $\{+g, \mathbf{0}_{z_k} -$

$g\}$ or $\{-g, \mathbf{0}_{z_k} + g\}$, where $+g$ and $-g$ cancel out, as is the case with pitch doubling and pitch halving errors.

- (b) subsequences of the form

$$\{m_l, \mathbf{0}_{z_{l+1}}, m_{l+1}, \mathbf{0}_{z_{l+2}}, m_{l+2}, \dots, \mathbf{0}_{z_{l+k}}, m_{l+k}\}$$

where $\sum_{l}^{l+k} m_i$ a music interval perceived by the human ear.

- (c) combinations of the above two cases

Other deviations can also be identified due to different performing styles of the instrument players.

In [5], the authors have dealt with the above deviations by adding states to a standard HMM following a strict modeling approach. This gave satisfactory recognition accuracy, at the expense of a significant increase in the number of states of the model.

In this paper a different approach is adopted. Instead of increasing the number of states, we focus on the Viterbi algorithm that finds the single best state sequence, $Q = q_1 q_2 \dots q_r$, for the given observation sequence $\mathbf{D} = \{d_1 d_2 \dots d_M\}$ in a variable duration HMM \mathcal{H} . Let us define the forward variable $a_t(j)$ as ([2])

$$a_t(j) = P(d_1 d_2 \dots d_t, \text{state } j \text{ ends at } t | \mathcal{H}), j = 1 \dots S \quad (1)$$

$a_t(j)$ stands for the probability that the model finds itself in the j -th state after the first t symbols have been emitted. It can be proved that ([2], [3])

$$a_t(j) = \max_{1 \leq \tau \leq T, 1 \leq i \leq S, i \neq j} [\delta_t(i, \tau, j)] \quad (2)$$

$$\delta_t(i, \tau, j) = a_{t-\tau}(i) A_{ij} p_j(\tau) \prod_{s=t-\tau+1}^t B_j(d_s) \quad (3)$$

where τ is the time duration variable, T is its maximum allowable value within any state, S is the total number of states, A is the state transition matrix, p_j is the duration probability distribution at state j and B is the symbol probability matrix. The overall recognition probability p^* is computed from

$$p^* = \max_{1 \leq j \leq S} a_M(j)$$

assuming a symbol sequence of length M . Equations (2) and (3) suggest that there exist $(S \times T - T)$ candidate arguments $\delta_t(i, \tau, j)$ for the maximization of each quantity $a_t(j)$. In order to retrieve the best state sequence, i.e., for backtracking purposes, the state that corresponds to the argument that maximizes equation (2) has to be stored in a two-dimensional array ψ , as $\psi(j, t)$.

Therefore,

$$\psi(j, t) = \arg \max [\delta_t(i, \tau, j)], 1 \leq \tau \leq T, 1 \leq i \leq S, i \neq j$$

In addition, the number of symbols actually spent on state j is stored in a two-dimensional matrix c , as $c(j, t)$.

3.3 The modification to the computation of the best state sequence

Let us now return to the deviations of type (a) and without loss of generality, let us focus on deviations of the form $\{+1, \mathbf{0}_{z_k} - 1\}$. As it was previously stated, $+1$ and -1 are not real music intervals and if they were canceled out by summation, the resulting subsequence would consist entirely of zeros, i.e., would be of the form $\mathbf{0}_{z_k+2}$. This suggests that if, in the original \mathbf{D} sequence, $\{+1, \mathbf{0}_{z_k} - 1\}$ was replaced by $\mathbf{0}_{z_k+2}$ the resulting $\hat{\mathbf{D}}$ sequence would give a non-zero recognition probability, whereas \mathbf{D} would result in a zero probability. Therefore, for the Z-states, it is desirable to modify equations (2) and (3), so as to reflect the need to be able to check for subsequences that contain symbols that can be canceled out. We notice that in equation (2), each candidate argument refers to τ symbols of the observation sequence and this is why the product in (3) $\prod_{s=t-\tau+1}^t B_j(d_s)$, is calculated. If the value of $\sum_{s=t-\tau+1}^t d_s$ is equal

to zero, this indicates a possible symbol cancellation. That is, one must also consider the possibility that a series of τ zeros have been emitted instead. This is quantified by considering $(S \times T - T)$ additional δ arguments to augment equation 2, namely

$$\hat{\delta}_t(i, \tau, j) = a_{t-\tau}(i) A_{ij} p_j(\tau) \prod_{s=t-\tau+1}^t B_j(d_s = 0) \quad (4)$$

It must be noted that equation (4) is computed only if $\sum_{s=t-d+1}^t d_s = 0$.

If, on the other hand, $a_t(j)$ refers to an S-state, then symbol cancellation is desirable if the resulting sum, $\sum_{s=t-\tau+1}^t d_s$ is equal to the actual symbol associated with the S-state. If this holds, the whole subsequence of symbols is treated as one symbol equal to their sum and again, $(S \times T - T)$ additional δ arguments must be computed for $a_t(j)$ according to the following equation:

$$\hat{\delta}_t(i, \tau, j) = a_{t-\tau}(i) A_{ij} p_j(\tau) B_j(\sum_{s=t-\tau+1}^t d_s) \quad (5)$$

This symbol cancellation allows the model to deal with deviations of type (b). Maximization is now computed over all δ and $\hat{\delta}$ quantities.

If, for some states of the best-state sequence a symbol cancellation took place, it is useful to store this information in a separate two-dimensional matrix s , by setting the respective $s(j, t)$ element equal to "1" (a zero indicates that no symbol cancellation took place). Matrices ψ and c are still used for backtracking purposes.

4. TRAINING THE HMMS

We have so far focused on the required modifications to the Viterbi algorithm in order to calculate the recognition probability and best state sequence of a symbol string \mathbf{D} given a HMM \mathcal{H} . We now turn our attention to the training algorithm for the HMMS. For each type of musical pattern, a set of K symbol sequences, $\mathbf{O} = [\mathbf{D}^{(1)}, \mathbf{D}^{(2)}, \dots, \mathbf{D}^{(K)}]$, is chosen for the training phase. The goal of the training stage is to adjust the model parameters to maximize the probability $P(\mathbf{O} | \mathcal{H}) = \prod_{l=1}^K P(\mathbf{D}^{(l)} | \mathcal{H})$. This is achieved by means of an iterative procedure. During each iteration, each observation sequence $\mathbf{D}^{(l)}$ is presented to the input of the respective variable duration HMM and for each $\mathbf{D}^{(l)}$, the best-state sequence is calculated by means of the modified Viterbi algorithm that was introduced in section 3.3. Let BP^l denote the best state sequence associated with symbol sequence \mathbf{D}^l . The reestimation equations for the Viterbi algorithm stem from the cumulative processing of all best-state sequences, namely

- $\hat{A}_{ij} = \frac{\text{no. of transitions from state } i \text{ to state } j}{\text{no. of transitions from state } i}$, $\forall i \neq j$, taking into account all BP^l 's. If $i = j$ then $A_{ij} = 0$ by the definition of the variable duration HMM.
- $\hat{\mu}_i = \text{average number of symbols spent at state } i$, taking into account all occurrences of state i in all BP^l 's.
- $\hat{\sigma}_i = \text{deviation from } \mu_i \text{ of the number of symbols spent at state } i$, taking into account all occurrences of state i in all BP^l 's.
- We have chosen not to reestimate the B matrix for reasons that will be explained below.

4.1 Special cases of the A matrix

Although, the above reestimation formulas are valid in the general case, for a large number of musical patterns that we studied, certain simplifications are possible for the reestimation of the state transition matrix A .

First of all, if we assume that all notes are present in every instance of the musical pattern, then in this case it suffices to initialize the state transition matrix with 1's on the first diagonal and no reestimation of A is needed. Equivalently, this means that each Z-state Z_i

in the best path is followed by an S-state S_i and each S_i is necessarily followed by Z_{i+1} .

Secondly, if we want to deal with missing notes, then state transitions $Z_i \rightarrow S_{i+1}$ should also be made possible. In the state transition matrix the respective elements $A(Z_i, S_{i+1})$ must be initialized with positive values. In this case, reestimation can still be omitted, if we are confident with the initialization values. If, however, we want to proceed with the reestimation of A , then, clearly, A is upper triangular and only values on the first and second diagonal will be affected.

A more complex case occurs, if a musical pattern consists of sub-patterns that repeat themselves in a row, with a varying number of repetitions. Such is the case with several musical patterns in the context of Greek Traditional music, due to the expressive performance of instrument players. In this case it is desirable to include backward transitions as well, upon initializing the HMM. Although, the A matrix will no longer be upper triangular, the reestimation formula still holds.

4.2 The symbol probability matrix B

We have assumed that only zeros are emitted from a Z-state and each S-state can only emit one symbol. Thus, each column of B has one element whose value is equal to 1, $B_{Z_i}(d_s = 0) = 1$ and all other elements are zero valued. For each S-state, the respective column has also one element whose value is equal to 1, $B_{S_i}(d_s = m_i) = 1$ and all other elements are zero valued. Therefore no reestimation of B is necessary. For the patterns that we studied, we observed that, due to the performance of instrument players, in a limited number of cases, certain music intervals can be one quarter tone higher or lower than expected. This can be accommodated if, each S-state is also allowed to emit, with a small probability, symbols that are one quarter tone higher/lower than the symbol associated with the state. Due to the limited number of such cases, reestimation can still be avoided.

5. APPLICATION OF THE METHOD IN THE CONTEXT OF GREEK TRADITIONAL MUSIC

The musical system of Greek Traditional music and the techniques of instrument players give the resulting sound material a radically different structure when compared with that of the western equal - tempered intervalic system (system of musical scales) [10],[5]. Although the proposed classification scheme is also applicable to western type of music, we chose Greek Traditional music since this poses more difficulties in recognition.

From a large number of types of transitory musical patterns, encountered in practice in different instrumental styles, we have selected the twelve most typical cases, in a monophonic environment. Here, the term monophonic refers to a single non-polyphonic instrument, such as the Greek Traditional clarinet, recorded under laboratory conditions with an ambient noise of less than 5dB. The Greek Traditional clarinet is an instrument that closely resembles the western-type clarinet.

The choice of the types of patterns was suggested by musicologists on the basis of a) their common use in practice and b) their respective time elasticity. The time elasticity of a musical pattern refers to the phenomenon of stretching its total length, up to five times in some cases, while retaining its musical function. It is assumed that the patterns to be classified have been isolated from their context by means of a manual segmentation process.

A set of 1200 musical patterns were generated by four professional Greek Traditional Clarinet players, involving all the aforementioned twelve types of musical patterns. For the feature generation stage, a number of robust fundamental frequency tracking algorithms were employed, including Brown's narrowed autocorrelation method [7], Tolonen's multipitch analysis model [8] and Brown's pattern recognition algorithm based on the constant-Q transform [9]. As an example, consider the narrowed autocorrelation method, where we used a moving window of 1024 samples (multiplied by a Hamming function) and narrowing was achieved by means of four

shifted versions of the autocorrelation function.

For the quantization step, quarter-tone resolution was adopted and an alphabet of 121 discrete symbols was used, implying music intervals in the range of $-60 \dots +60$ quarter-tones, i.e. $G = 60$.

The training set for each musical pattern consisted of 60 observation sequences, that exhibited the deviations from the prototype presented in section 3.2. A total of 600 patterns were used as the test set.

At an average, for each variable duration HMM, the reestimation algorithm converged after 40 iterations. Parameter T , which defines the maximum allowable time spent at each state, was set equal to 100 symbols for the Z-states and 20 symbols for the S-states. This was due to the time elasticity of the musical patterns that we studied.

Our tests were carried out using Matlab. The overall recognition rate was above 95%.

6. CONCLUSIONS AND FUTURE RESEARCH

This paper provides a modified Viterbi algorithm for the calculation of recognition probabilities generated by variable duration Hidden Markov models, capable of providing increased classification accuracy of musical patterns to predefined classes. A methodology for constructing such HMMs is also presented and various issues related to the training algorithm of the HMMs are analyzed. A case study involving patterns originating in the context of Greek Traditional music is also presented. The suggested approach can be useful in a number of content-based retrieval systems including query-by-humming and repeated pattern finding systems. In the future, the approach will be examined in relation with continuous observation HMMs and segment models, permitting to depart from the discrete observation hypothesis. Multi-dimensional feature vectors will be studied including timbral and rhythmic features.

REFERENCES

- [1] B. Pardo, W. Birmingham, "Query by Humming: How good can it get?", *Proc. of the Workshop on Music Information Retrieval, SIGIR 2003*, Toronto, Canada, 2003.
- [2] L.R. Rabiner, "A Tutorial on Hidden Markov Models and Selected Applications in Speech Recognition", *Proceedings of the IEEE*, Vol. 77, No. 2, 1989
- [3] J. Picone, "Continuous Speech Recognition Using Hidden Markov Models", *IEEE ASSP MAGAZINE*, July 1990.
- [4] S. Theodoridis, K. Koutroubas, "*Pattern Recognition*", 2nd Edition, Academic Press, 2003.
- [5] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns using Hidden Markov Models", *Lecture Notes in Computer Science and Artificial Intelligence*, LNCS/LNAI 2445, Springer Verlag, pp. 133-143, 2002.
- [6] A. Pikrakis, S. Theodoridis, D. Kamarotos, "Recognition of Isolated Musical Patterns using Context Dependent Dynamic Time Warping", *IEEE Transactions on Speech and Audio Processing*, vol. 11(3), pp. 175-183, 2003.
- [7] J.C. Brown and B. Zhang, "Musical frequency tracking using the methods of conventional and narrowed autocorrelation", *Journal of the Acoustical Society of America*, Vol. 89, No 5, 1991
- [8] T. Tolonen and M. Karjalainen, "A Computationally Efficient Multipitch Analysis Model", *IEEE Transactions on Speech and Audio Processing*, Vol. 8, No 6, November 2000
- [9] J.C. Brown, "Musical fundamental frequency tracking using a pattern recognition method", *Journal of the Acoustical Society of America*, vol. 92, No 3, 1992
- [10] S. Karas, *Theoritikon - Methodos on Greek Traditional Music*, ed. Athens, 1982 (in Greek)