Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 46 (2015) 381 – 387

Procedia Computer Science

International Conference on Information and Communication Technologies (ICICT 2014)

Music Information Retrieval for Polyphonic Signals using Hidden Markov Model

S Chithra, M S Sinith*, A Gayathri

Government Engineering College Thrissur, Kerala, India-680 009

Abstract

Now-a-days, almost all music can be easily accessed via the Internet, but at the same time music can be hard to find. This has created the demand for intelligent music retrieval which allows the user to access the songs that he or she likes. The idea of music information retrieval is basically used in music search systems. In a music search system there will be a huge database of songs. For an efficient music search system, when a particular song in the database is requested, the song has to be correctly identified and retrieved from the database. Music information retrieval for polyphonic music is presented here.

Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)

Keywords: Hidden Markov Model; Music Information Retrieval.

1. Introduction

Most of the music information search system is limited to monophonic songs or polyphonic songs in a particular language or a particular musical category. The monophonic songs consists of single musical note at a time while the polyphonic songs consists of more than one musical notes at a time but being perceptible as one total piece of music. In the proposed music retrieval system, the song selection is not restricted to monophonic songs or to the language or the musical category of the song. Polyphonic music is more complicated than a monophonic music. An example for a polyphonic music is film songs in which one or more singers are sing along with different musical instruments. For testing the algorithm we have selected songs in four different languages, English, Hindi, Malayalam and Tamil. Also in the database we have included all kinds of songs, which fall in different musical categories like classical music, jazz, rock etc.

The basic of music is the musical notes. According to South Indian Classical Music there are seven notes in music called the saptha swaras, the seven notes are Sa, Ri, Ga, Ma, Pa, Da, Ni. Some of the notes can again be divided and form a twelve note system and a twenty two note system. We are following the twelve note system which is used by R.Sridhar et.al. [1]. Except the two notes Sa and Pa all other notes can be again divided and form two notes, thus from

* Corresponding author. Tel.: +919847381870.
E-mail address: sinith@ieee.org

1877-0509
Peer-review under responsibility of organizing committee of the International Conference on Information and Communication Technologies (ICICT 2014)
doi:10.1016/j.procs.2015.02.034

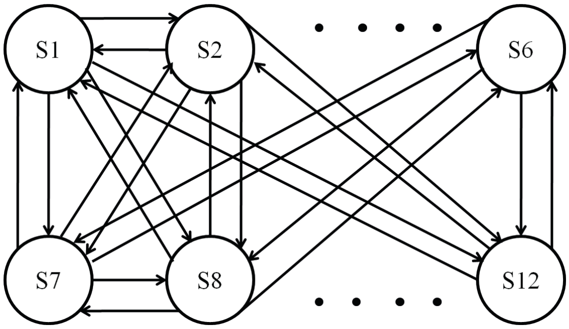

Fig. 1. HMM model used.

the seven note system we can form a twelve note system. The twelve notes are Sa, Ri1, Ri2, Ga1, Ga2, Ma1, Ma2, Pa, Da1, Da2,Ni1, Ni2, which are equally separated in the log-frequency domain.

A song can be represented using these twelve musical notes, it is not necessary that in a single song all the notes are present. A song can be considered as a linear combination of the notes. For a song, the first note can be Pa the second note can be Ni2 and the third note can be Ma1, for another song the first note can be Ma1 the second note can be Ni2 and the third note can be Pa, and all other notes could be absent. The two songs have the same notes present in them but the order of appearance of notes is different. Such a small variation results in two different songs. So using these twelve notes we can have infinite number of songs.

The twelve notes have a distinct fundamental frequency, so the notes can be identified by using their fundamental frequencies. There are several fundamental frequency tracking algorithm, both in frequency and in time domain. One of the frequency domain method is Schroeders histogram discovered by M. R. Schroeder [2]. Modified Schroeders histogram is used in the work.

2. Hidden Markov model

Hidden Markov Models (HMM) is a doubly stochastic process, in which one of the stochastic process is not observable or hidden and other stochastic process is observable. The challenge in Hidden Markov model is to find out the hidden stochastic process from the observable stochastic process. The elements of the hidden stochastic process is called the states and the elements of the observable stochastic process is called the observations. The states and the observations are related as, at a particular time instant an observation can be generated from the state with a certain probability called emission probability. In the next instant another observation can be generated from another state. Each time only the observation is visible to the observer, the states are not visible to the observer[3].

Consider N states $\{S_1,S_2,......S_N\}$ and let us denote the set of all possible output symbols as $V = \{v_1,v_2,.....v_M\}$ the output symbol at time t as O_t. The sequence of observed symbols is denoted as $\{O_1,O_2,.....O_T\}$. The HMM can be completely defined by three parameters, i.e, the transition probability matrix A, the emission probability matrix B, and initial probability matrix π , as introduced by L. R.Rabiner[4].

$$\lambda = (A, B, \pi) \tag{1}$$

Where A is the State transition matrix, $A=\{a_{ij}\}$. The element a_{ij} is the probability of transition from state S_i to state S_j. B is the Emission probability matrix, $B=\{b_{ij}\}$, b_{ij} is the probability that out put O_j comes from the state S_i. The initial probability matrix $\pi=\{\pi_i\}$, where π_i is the probability that S_i is the starting state.

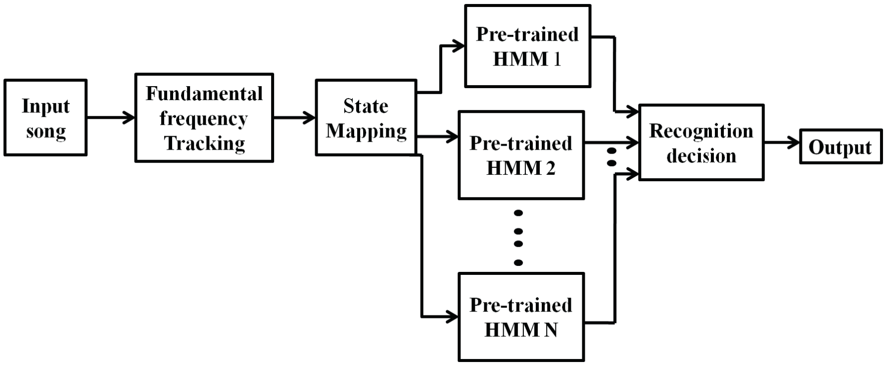

Fig. 2. Music search system.

2.1. Hidden Markov Model used

Here the musical notes are taken as the states as taken by Gaurav et al.[5]. The HMM used in the algorithm consists of 12 states, i.e, the 12 musical notes. The observations are the fundamental frequency of the notes. Here from one state only one observation is generating. So the emission probabilities have only two possible values, one and zero.

$$b_{ij} = 0, \quad \forall i \neq j \tag{2}$$

$$b_{ij} = 1, \quad \forall i = j \tag{3}$$

The HMM model used in the algorithm is shown in the fig. 1. It consists of 12 states represented as state 1 to state 12. State 1 corresponds to the first musical note Sa, state 2 corresponds to second musical note Ri1 and so on. The states are related to each other with certain transition probabilities. Since there is a one to one relation between the musical notes and their fundamental frequencies there is no need to show the relation between the states and the observations in the HMM mode.

3. Musical search system

The block diagram of the proposed music information retrieval system is shown in fig. 2. Suppose there are N songs in the database and we generate a unique HMM for each song, these N HMMs will be stored in the database. There are six blocks in the block diagram, ie, the input song, fundamental frequency tracking, state mapping, the pre trained HMMs, recognition decision and the output.

When a song is requested to be retrieved from the database that song is taken as the input song. Since the songs in the database are stored as their HMM and the base of a HMM is the states, for the identification process the states of the input song has to be identified. The two steps, fundamental frequency tracking and the state mapping together will find out the states present in the input song. After finding out the states of the input song it has to be compared with the N HMMs present in the database. In the decision making block the HMM which give maximum matching with the input song is decided as the requested song and finally given as output. The output of an efficient music information retrieval system is expected as the requested song itself.

4. Feature generation

The music information retrieval system has two sections, first one is the HMM generation and the second one is the musical search. For both of the sections the steps to find out the notes are same. To find out the notes present in a song there are three steps, first step is the fundamental frequency tracking, the second step is the quantization of obtained fundamental frequency and the third step is the mapping of quantized fundamental frequencies to states of HMM.

4.1. Notes in a song

The first step is to find out the notes present in the song. The size of a note is taken as 1024 samples as taken by M.S. Sinith and K. Rajeev [6]. So the song is divided into frames by using a window size of 1024 samples for the song of sampling frequency 44100 Hz. The step increment is taken as 1024 samples, so there is no overlapping between the frames. For each note the fundamental frequency present is found. For the twelve musical notes there are distinct fundamental frequencies. So by finding out the fundamental frequency we can finally determine the note present in the frame.

4.2. Fundamental frequency tracking algorithm

The fundamental frequency of the notes are taken as the observations in the HMM. A frequency-domain algorithm called Schroeders histogram is used for the fundamental frequency tracking. In the Schroeders histogram, the missing of fundamental frequencies are not considered. So to avoid that a modification is made in the Schroeders histogram.

The modified Schroeders histogram is applied to each and every frame of the song. The first step of modified Schroeders histogram is finding out the fast fourier transform of the note. We take a 1024 time fft of the frame of size 1024. The next step is to find out the magnitude spectrum. The magnitude spectrum will be symmetric with respect to the index point 512. So only the first 512 values of the magnitude spectrum is considered. The fundamental frequency is taken as the frequency at which the magnitude spectrum has maximum value as taken by Judit and Bin [7]. In order to avoid the missing of fundamental frequency we take the first ten higher values in the magnitude spectrum and their indices, and Greatest Common Divisor(GCD) of each of each and every pairs of them are taken. Mode of this is taken to obtain the most recurring value. Thus we will obtain the fundamental frequency of the frames in a song.

4.3. Quantization of extracted frequencies

The extracted fundamental frequency is passed to a nonuniform quantizer having twelve levels. The quantization step is $24*\log(\text{JI})$ which is derived by Arvindh Krishnaswamy [8]. Where JI is Just Intonation ratio, Just Intonation is the ratio of fundamental frequency f_i to the frequency of C note, called sruthi,in South Indian classical music.

4.4. Mapping frequencies to states of Hidden Markov Model

The twelve quantized frequencies so obtained are mapped to twelve states of the hidden markov model which are S_1,S_2,S_3,...S_{12}.

5. HMM generation

After finding out the sequence of fundamental frequencies or states present in a song we have to find out the unique HMM of the song such that it should have the high probability of generating the sequence of fundamental frequencies or states of the song.

There are three basic problems in HMM, evaluation problem, decoding problem and learning problem, it is explained by B. H. Juang and L. R.Rabiner [9]. Out of which we are facing the learning problem, i.e, for a Hidden Markov model, a set of observed sequences are given, here the problem is to find what should the model parameters be so that it has a high probability of generating those sequences. To solve this problem we are following the Baum-Welch algorithm.

The Baum-Welch algorithm is a particular case of a generalized expectation-maximization (GEM) algorithm. The Baum-Welch algorithm is used to find the unknown parameters of a hidden Markov model (HMM).

Since there is a one to one relation between the states and observations the emission probability matrix is always an identity matrix and we are not considering the initial state, so there is no need to consider the emission probability matrix and initial state matrix. Then the only changing parameter is the transition probability matrix. So out of the three parameters of HMM, we are considering only the transition probability matrix A. So HMM of a song means only the transition probability matrix A, i.e, HMM of a song is a 12X12 matrix. Initial transition probabilities are taken as equiprobable, i.e, a_{ij} = 1/12. Then according to the states of the song HMM of the song is trained or finding out new value for transition probability matrix, using Baum-Welch algorithm.

According to the Baum-Welch algorithm we will find out two parameters γ_t(i) and ξ_t(i,j) Where γ_t(i) is the probability of being in state S_i at time t, given the observation sequence O, and the model λ, explained by Han et al.[10].

$$\gamma_t(i) = P[S = S_i|O, \lambda] \tag{4}$$

The second parameter ξ_t(i,j) is the probability of being in state S_i at time t, and state S_j at time t + 1, given the observation sequence O, and the model λ.

$$\xi_t(i, j) = P[S_t = S_i, S_{t+1} = S_j|O, \lambda] \tag{5}$$

From the two parameters we will find out the new value for the transition probability a_{ij}.

$$a_{ij} = \sum_{t=1}^{t=T-1} \xi_t(i, j) / \sum_{t=1}^{t=T-1} \gamma_t(i) \tag{6}$$

The numerator is the expected number of transitions from state S_i to S_j. The denominator is expected number of times state S_i occurred.

6. Musical search

When a song is given as input the musical search system has to retrieve a song from the database, for an efficient musical search system the input song itself retrieved from the database. The feature used for the recognition or retrieval process is the summation of the probability emitted by the HMM. The model producing the highest value is recognized to be the pattern of the wave file given as input.

7. Result

In the database we have included polyphonic music. Film songs are the best examples of polyphonic music, so we have taken film songs in different languages, i.e, songs in Malayalam, Tamil, Hindi and English. Also we have selected songs of different musical categories like classical music, jazz, rock etc.

A film song has a duration of about 5 minutes, if we take the full length of the song the running time will increase as the size of the database increases. So for the testing purpose we take the songs with duration about 25 seconds. For that after reading the full song we take the first 5 lakh points of the song for generating HMM and testing the algorithm.

We tested the polyphonic music in five steps and for testing the algorithm we have already generated the HMMs for all the songs and stored in the database. In the first step we consider a database with only twenty songs and we tested the algorithm by giving each and every song in the database as input song and obtained the output. Then we repeat the same process for the next four times by increasing the size of database by new twenty songs and the results are given in figure 3.

In the first step we have included twenty songs in the database. Out of twenty songs the algorithm is correct for eighteen songs and wrong for only two songs, i.e, the algorithm is correct for 90 percentage. In the second step we have extended the size of the database to forty songs. Out of forty songs the algorithm is correct for thirty seven songs and wrong for only three songs, i.e, the algorithm is correct for 92.5 percentage. In the third step we have extended the size of the database to sixty songs. Out of sixty songs the algorithm is correct for fifty five songs and wrong for only five songs, i.e, the algorithm is correct for 91.6 percentage. In the fourth step we have extended the size of the database to eighty songs. Out of eighty songs the algorithm is correct for seventy three songs and wrong for only seven songs, i.e, the algorithm is correct for 91.25 percentage. In the fifth step we have extended the size of the database to hundred songs. Out of hundred songs ninety two songs are correct and only eight songs are wrong. The correctness of the algorithm is 92 percentage.

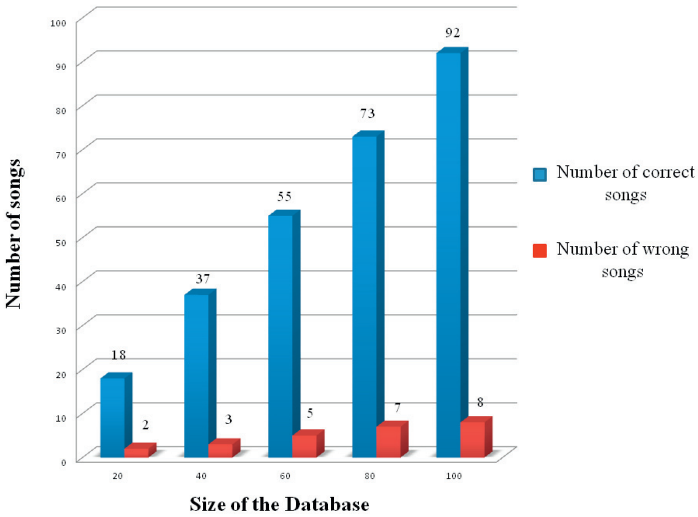

Fig. 3. Result.

Table1 gives the details about the wrongly identified songs. When the size of the database is hundred there are eight wrongly identified songs and those are listed in the table 1. The database includes 25 Malayalam songs, 25 Hindi songs, 25 Tamil songs and 25 English songs. The songs are of different musical categories like Rock, Jazz, Indian rock, Indo jazz etc.

The main instruments that are used in Rock music are electric guitar, drums, piano and keyboard. In the case of Jazz music the main instruments are guitar, drums, piano, saxophone, clarinet etc. Indian rock music is a musical genre that incorporates Indian music with main stream rock music. Indo jazz music consists of Indian classical instruments like violin, sitar, tabla etc.

Table 1. Category and rank of wrongly identified songs.

Sl.No	Type of the input Song (*t*)	Type of the wrongly identified song (*t*)	Rank of the song(*t*)
1	Indo jazz	Jazz	2
2	Rock	Indian rock	3
3	Jazz	Jazz	2
4	Indian rock	Indo jazz	2
5	Rock	Indo jazz	2
6	Indian rock	Indian rock	2
7	Rock	Indo jazz	3
8	Indian rock	Indo jazz	5

From table 1 we can see that for 50 percentage cases the wrongly identified songs are also in the same musical category as the input song. Since the songs of same musical category shares the same instruments and mood, there can be a similarity between the songs and so there is chance of similarity between their HMMs. Since Indian rock and Indo jazz are basically Indian music, in some cases an Indian rock music can be more similar to an Indo jazz music or vice versa than to a Rock music or to a Jazz music.

The last column of table 1 is the rank of the song. Rank of the song means, if the HMM of a particular song gives maximum matching to that song, then the rank of the song is one. When the HMM of the song gives only second matching to the same song then the rank of the song is two and so on. Rank of the song indicates how well the HMM can represent a song or we can take it as the strength of the HMM. As the value of the rank of song increases the strength of the HMM decreases.

From table 1 we can see that rank of the wrongly identified songs are less than or equal to five out of hundred. In most of the cases the rank is two, it means that the HMM generated has still a comparative strength, i.e, if the song which is wrongly identified as the input song is not present in the selected database then the actual input song will be correctly identified.

8. Conclusion

The results obtained for Music information retrieval for polyphonic music is promising one. The number of polyphonic songs for which the algorithm did not work is small compared the number of correct songs. The correctness of the algorithm for polyphonic songs in all five step is always greater than 90 percentage.

We can also say that the HMM generated using the algorithm has high strength, because out of 100 HMMs 92 HMMs have a rank equal to one. For the 8 wrong cases the rank of HMM is less than or equal to 5.

References

1. Rajeswari Sridhar, T.V. Geetha. Swara identication for south indian classical music. *ICIT'06*. 2006.
2. M. R. Schroeder.Period histogram and product spectrum: New methods for fundamental-frequency measurement. *Journal of the Acoustical Society of America*. 1968.
3. Przemyslaw Dymarski. *Hidden Markov Models, Theory and applications*. Croatia:InTech; 2011.
4. L. R. .Rabiner. A tutorial on hidden markov model and selected application in speech recognition. *Proceedings of IEEE*. 1989.
5. Gaurav Pandey, Chaitanya Mishra,Paul Ipe. Tansen : A system for automatic raga identication. *Indian International conference on Articial Intelligence*. 2003.
6. M.S. Sinith, K. Rajeev. Pattern recognition in south Indian classical music using a hybrid of HMM and DTW. *ICCIMA*. 2007.
7. Judith C . Brown, Bin Zhanga. Musical frequency tracking using the methods of conventional and "narrowed" autocorrelation. *Acoustical Society of America*. 1991.
8. Arvindh Krishnaswamy. Application of pitch tracking to south indian classical music. *ICASSP*. 2003.
9. B. H. Juang, L. R. Rabiner. An introduction to hidden markov models. *IEEE ASSP MAGAZINE*. 1989.
10. Han Shu, I. Lee Hetherington, James Glass. Baum-Welch training for segment-based speech recognition. *ASRU* . 2003.