CHAPTER 6

USE OF THE CICC FOR THE IDENTIFICATION OF NON-MUSIC COMPONENTS

As discussed in Chapter 5, features such as Tonic and Carnatic Interval Cepstral Coefficients were extracted, and used by the Raga model for the identification of the Raga. In this Chapter we discuss the use of these features for the extraction of non-music information contained in the music signal, such as Singer, Instrument, Emotion, and Genre. As discussed in Chapter 2, the Singer, Instrument, Genre and Emotion have been identified by extracting Spectral features like the spectral flux, spectral centroid, autocorrelation and Cepstral features, like the MFCC, and OFCC, and constructing a GMM, HMM or SVM using these feature values (Kim and Whitman 2002, Nwe and Li 2007, Panagakis et al 2008, Li and Ogihara 2004, Yang et al 2006).

6.1 SINGER IDENTIFICATION

The task of automatic Singer identification has been proposed for Western music by researchers. In one of the methods (Kim and Whitman 2002) voice coding features were used for Singer identification. They have used acoustic features derived from LP coefficients and formant frequency, for the construction of the Gaussian Mixture Model (GMM) and the Support Vector Machine (SVM), and have compared the results of the performance of these models. The performance of the SVM for Singer identification was better, when the input had a voice or a combination of voice and non-voice.
On the other hand, the GMM performed well, only if the input was a monophonic voice signal.

Another work on Singer identification uses the vibrato characteristics of the Singer which are essentially periodic fluctuations in pitch of the Singer (Nwe and Li 2007). The vibrato characteristics are predominant in a singing voice, and are reflected in the coefficients that are extracted. The authors have used the Octave Frequency Cepstral coefficients (OSCC/OFCC), which were proposed by Maddage et al (2004). As already discussed, the OSCC / OFCC are based on the Octave scale of the keyboard, where the cut-off frequency of each of the filter banks is based on the frequency assigned to the keys of an octave. The authors have extracted 9 coefficients, and used them for constructing a Hidden Markov Model (HMM), which is used as the model space for Singer identification.

In another technique for Singer identification (Mesaros et al 2007, Levy 1982), the authors have used the MFCC with a new distance measure to perform Singer identification.

In general the process of Singer identification for Western music was based on identifying the characteristics conveyed, using the MFCC and OSCC features.

The techniques proposed for Western music Singer identification based on the OSCC, cannot be directly used for Carnatic music Singer identification, because of the fundamental difference in the music characteristics of the two systems, as already discussed in Chapter 1 (Sambamurthy 1983).

In addition, the techniques related to the Octave Frequency Cepstral coefficients are based on Western music’s major scale frequency band, and
cannot be directly used for Carnatic music. The MFCC were devised for Speech processing and in general, their use for Music signal processing is not justifiable. However Cepstral Coefficients convey the timbral characteristics, and are therefore, suitable for Singer identification. Hence, in this work, we use our specially designed Carnatic Interval Cepstral Coefficients (CICC) for identifying the Singer by constructing the GMM to form the basis for Singer identification. The concept of LDA discussed earlier has also been tried for identification of the Singer.

6.1.1 GMM for Singer Identification

After extracting the Carnatic Interval Cepstral coefficients, they are used to construct a Gaussian mixture model (GMM). The Gaussian mixture architecture estimates the probability density function for each class, and then performs the classification by applying the Baye’s theorem (Pawlak 2002).

We use the expectation maximization algorithm to construct the GMM using the Carnatic Interval Coefficients that were extracted from the various segments of the input voice only signal. The Gaussian Mixture model is constructed for every Singer, and the mean and covariance of the various clusters are estimated. These values of the mean and covariance are used to represent the Singers during the training phase.

For the process of testing the Singer, the same set of Carnatic Interval Cepstral Coefficients is extracted to determine the mean and covariance. The extracted mean and covariance are compared with the existing clusters, using a distance measure. In this work, we have used the Mahalanobis distance measure to estimate the distance between the clusters, to find the similarity to identify the Singers (Xiang et al 2008). From the literature it has been observed that this distance measure is well suited for music signal processing (Mesaros et al 2007). It differs from the Euclidean
distance, in that it takes into account the correlations of the data set, and is
scale-invariant, i.e., not dependent on the scale of measurements.

Formally, the Mahalanobis distance from a group of values with
mean \( \mu = (\mu_1, \mu_2, \mu_3, \ldots, \mu_p)^T \) and covariance matrix \( \Sigma \) for a multivariate vector
\( x = (x_1, x_2, x_3, \ldots, x_p)^T \) is defined as in Equation (6.1)

\[
D_M(x) = \sqrt{(x - \mu)^T \Sigma^{-1}(x - \mu)}.
\]  

(6.1)

The Mahalanobis distance can also be defined as the dissimilarity measure
between two random vectors \( \bar{x} \) and \( \bar{y} \) of the same distribution with the
covariance matrix \( \mathbf{P} \) as given in Equation (6.2)

\[
d(\bar{x}, \bar{y}) = \sqrt{\mathbf{P}^{-1}(\bar{x} - \bar{y})^T(\bar{x} - \bar{y})}.
\]  

(6.2)

If the covariance matrix is the identity matrix, the Mahalanobis distance
reduces to the Euclidean distance. If the covariance matrix is diagonal, then
the resulting distance measure is called the normalized Euclidean distance.
This distance measure is used as a baseline, and is based on determining the
minimum distance value between the training phase and the testing phase,
where the Singer is identified.

6.1.2 Comparison of Performance between MFCC and CICC - GMM

The input music signal is sampled at a rate of 44.1 KHz. As
explained in Chapter 4, the input voice only signal is segmented, and then,
features including the Carnatic Interval Cepstral Coefficients (CICC) and the
Mel-frequency Cepstral Coefficients (MFCC) are extracted from the
individual segments of the signal.

The coefficients are then used to construct the GMM. Using the
constructed GMM, the mean and covariance of the various clusters are
determined. The GMM model is constructed for approximately 15 Singers, by
training it with totally 1000 songs, 60 songs per Singer. The songs that we have considered include traditional Carnatic songs and songs from Tamil movies based on Carnatic Ragas. After constructing the model we use Mahalanobis distance measure, and identify the Singer from a total of 1500 songs belonging to different Singers.

The same experiment is carried out by estimating the MFCC as features instead of the CICC, with other procedures remaining the same. It has been observed that the CICC gave better results compared to the MFCC, especially in the case of the identification of a new Singer. The algorithm that uses the CICC identified the input with a new Singer as belonging to a new cluster, while the algorithm that uses the MFCC tried to match the input with the existing clusters.

The graph showing the performance is plotted in Figure 6.1.

Figure 6.1 Comparison of MFCC vs CICC for Singer identification using GMM
As can be seen from Figure 6.1, the incorporation of the Tonic in the CICC features helped to convey the timbral characteristics better, and thereby increased the identification rate. This is evident for Singers whose fundamental frequency has a considerably wider range, namely, Nithyasree, M.S. Subbulakshmi and K.J. Yesudas, where the CICC algorithm shows an increase of 25% in the performance. There is also an increase in the performance, when musicians sing for Tamil movies, where the Singer is expected to change the starting frequency depending on the song, in order to convey the Emotion and hence, sing with a wider range. The use of the MFCC features was not able to identify the Singer Sowmya, due to her vibrato characteristics, and in fact, her songs were mapped with the clusters belonging to other Singers. On the other hand, the use of the tonic in the CICC helped to identify Sowmya with an identification rate of 20%. For other Singers who use almost the same tonic for singing, the increase in the percentage of identification is not so marked.

6.1.3 LDA for Singer Identification

As discussed earlier in Chapters 2 and 5, the LDA has been used for determining music similarity (Hoffman et al 2008). In this work, we attempted Singer identification based on the construction of the Latent Dirichlet Allocation. In this work, the construction of the LDA is done using the extracted CICC features, which are used for defining the dirichlet parameters $\alpha$ and $\theta$. This approach is similar to the approach used for Raga identification, where the characteristic phrase was defined by the Raga Lakshana, while here the CICC were used as features. The parameters $\alpha$ and $\theta$ are represented using the mean and covariance of the CICC values. The generic parameter $\alpha$ is derived from the CICC values, by considering a
random mixture of Singers, while on the other hand, the value of $\theta$ is the
typical mean and covariance values for a particular Singer.

The LDA approach works as given below. Using the CICC, an
initial value for $\alpha$ is assumed by considering all Singers. Then a range of
songs is chosen sung by one Singer, and the probability value is computed
based on Baye’s theorem, to determine the individual Singer’s dirichlet
parameter $\theta$ based on the dirichlet parameter $\alpha$. From the values of the mean,
and covariance used in $\theta$, the value of $\alpha$ is recomputed, and this computation
is repeated until a converging value of $\theta$ is obtained for a single Singer. This
is the final model value of the LDA, which is represented in terms of the
parameters $\alpha$ and $\theta$, and is characteristic to the Singer. During the testing
phase the same CICC values are extracted, the value of $\theta$ is determined and
compared with the LDA model, to identify the probability of the Singer of the
input. The Singer whose parameter has the maximum probability value is
chosen as the Singer of the input music.

6.1.4 Comparison of Performance between MFCC AND CICC - LDA

For comparison purposes, the LDA for Singer identification is
carried out using the MFCC features instead of the CICC. The data set that
was used for the GMM is used for the construction of the LDA also. Based on
the Singer identification that was carried out, the performance of the LDA
using these two coefficients is shown in Figure 6.2.
The identification performances of the CICC are higher for almost all the Singers or the same as that of the MFCC. This is due to the incorporation of the Tonic in the computation of the CICC, which improves identification efficiency. However, the LDA model performed better than GMM model, since the LDA considered random mixtures of Singers, and hence, avoided the False positive and True negative error.

6.1.5 Comparison of GMM and LDA using MFCC and CICC

The graph given below in Figure 6.3 compares the performance of the GMM and the LDA for the process of Singer identification using the CICC and MFCC features respectively. As can be seen from the figure, the performance of the LDA is higher than that of GMM for all the Singers, except Nithyasree and M.S. Subbulakshmi. The variation is due to the reason
already discussed, which is the varying range of the tonic that these two Singers have during the performance. However, there are other drawbacks, in terms of incorrect computation of the tonic, yielding to error in the CICC values, the segments that have been chosen could be error prone, or noise could be present in the sample data. However, it can be concluded that, in general, the performance of the LDA for Singer identification using the CICC is better than that of the other algorithms.

Figure 6.3 Performance of the models GMM vs LDA using the coefficients MFCC / CICC for Singer identification

6.1.6 GMM for Duet-Singer Identification

The concept of a Duet is typical in the case of music from Tamil language movies. We tried the GMM model to identify the presence of duet Singers in a given musical piece and identify them. The training data that we used to construct the GMM is based on a solo Singer. In this thesis, we have not considered the Singers who sing in a chorus or simultaneously. However,
this work identifies the presence of Singers when they sing in an alternating fashion.

The proposed algorithm first identifies the presence of duet Singers in a given musical piece by determining the variation in the mean and covariance of the CICC values of the input signal at successive segments. If this variation between successive segments is predominantly high or low, then the presence of another duet Singer is identified. After identifying the presence, the values of the mean and covariance are compared with the GMM constructed for the individual Singers, and the Singers available in the input song are determined.

For the performance of Duet Singers nearly 500 songs from Tamil film music were considered, belonging to the combinations of P. Suseela & T.M. Sounderrajan, S.Janaki and S.P. Balasubramanian, S.P. Balasubramanian and Malaysia Vasudevan, K.J. Yesudas and S.Janaki, Mano and S. Janaki, and S.P. Balasubramanian and Chitra.

6.1.7 Comparison of Performance between MFCC and CICC

The performance of the algorithm is verified by means of the MFCC and CICC as features for the construction of the GMM. The corpus that was taken for the identification of duet Singers is from movies in Tamil language, and the results are shown in Figure 6.4. In this experiment also the CICC outperformed the MFCC for reasons already discussed.
In this work, if the algorithm identified both the Singers correctly, the performance is considered as positive, and even if one of the Singers is not identified the result was termed as negative. In this thesis, for the purpose of duet Singer identification we considered only Tamil Movie songs. Typically, the system wrongly identified songs sung by S.P. Balasubramanian and Mano. This is due to the fact that playback Singers especially, S.P. Balasubramanian and Mano, when they sing for a particular actor in a movie, have a tendency to adopt a tonic closer to the fundamental frequency of the actor, and moreover, they both have similar vibrato characteristics. Therefore, the songs sung by S.P. Balasubramanian and Mano in this Genre magnified the ambiguity in identification. In this algorithm also, the performance of the CICC is better than that of the MFCC. Carnatic music songs are not considered here, since in Carnatic music the Singers typically sing together simultaneously, which is not tackled by our techniques.
6.2 INSTRUMENT IDENTIFICATION

After identifying the Singer from the voice-only component of the input, the non-voice component is used for determining the Instrument present in the input signal. Martin and Kim (1998) concluded that identifying cue phrases in musical signals could result in better Instrument identification.

As discussed in Chapter 2, Instrument recognition was also carried out by isolating monophonic musical Instrument sounds using six features, namely, Cepstral coefficients, constant $Q$ transform frequency spectra, multidimensional scaling analysis trajectories, RMS amplitude envelopes, spectral centroid and vibrato. Kitahara et al (2003) exploited the pitch dependency on Timbre characteristics to improve the instrument recognition accuracy by 75%. This motivated us to try our Carnatic Interval Cepstral coefficients, which incorporated the tonic for Instrument Identification.

6.2.1 CICC and GMM for Instrument Identification

As discussed in Chapter 2, generally for Instrument Recognition, a model is constructed using spectral, and Cepstral features. In addition, the tonic also has been used as an additional feature to identify the Instruments. The work that has been done is mostly for identifying Instruments from a solo Instrument only signal.

As already explained, the CICC is a Cepstral feature that incorporates Tonic and conveys timbral characteristics, thereby making it a good choice for identifying Instruments. During the training phase, 800 songs were used. From the non-voice component of the signal separation algorithm, the tonic is determined using our algorithm corresponding to the middle octave ‘S’, segmented using the Tala based segmentation algorithm, and then
CICC features are extracted. These coefficients were used to construct the Gaussian mixture model.

In the testing phase, the comparison between the input signal’s features and the GMM is performed to identify the Instruments. For testing, we chose songs from Tamil film music and Carnatic Classical music which include a total of 1500 songs. The Instruments that we considered are the Violin, Veena, Flute and Tabla. These four Instruments included the class of string, wind and percussion Instrument families.

6.2.2 Analysis of Instruments Identified

We performed a comparison between the features of the MFCC and the CICC for the sample data, and analyzed the performance. The graph is plotted in Figure 6.5.

![Figure 6.5 Comparison of MFCC vs CICC for Instrument Identification](image-url)
The average identification rate of Instruments using the CICC features is 92%, which is higher than those using MFCC features, which is 88%. This is due to the fact that the CICC features incorporated the tonic component also. Therefore, the timbral features, which are typically used to identify the Instruments, have also considered the tonic characteristics. As claimed by Kitahara et al (2003), the inclusion of fundamental frequency makes the Carnatic Interval Cepstral Coefficients robust for Instrument identification.

6.3 EMOTION IDENTIFICATION

Emotion is yet another important component that is conveyed by a given musical piece, and can be deciphered even by a person who does not have musical knowledge.

Many of the algorithms that are available for recognizing Emotion in music have been derived from speech processing systems. These algorithms, when applied to music, try to look for speech specific characteristics rather than music characteristics. Researchers in Western music scenario have explored Emotion identification in music, by incorporating music characteristics like rhythm, tempo etc. In our work, we have adapted the algorithms used for speech and for Western music and incorporated some features of Carnatic music like Raga and tonic, for the process of Emotion identification. The basis of using Raga and tonic for Emotion detection is derived from the fact, that Carnatic music literature outlines the use of specific Ragas to convey specific Emotions.

As discussed in Chapter 2, using the features like the fundamental frequency, the mean, standard deviation of the signal, minimum and maximum peak frequency, speaking rate and up and down slopes of the frequency, and articulatory features (Dellaert et al 1996) Emotion is
recognized from speech. In addition to this, other features like the MFCC are used for constructing GMM to identify the Emotion from speech (Tang et al 2009).

Although the characteristics of music are different from those of speech, the MFCC was used to identify the Emotion from music (Li and Ogihara 2004). Yang et al (2006) have done work on identifying and classifying the Emotion of Western music, are based on Thayer’s 2D model (1989).

6.3.1 Emotion Classification in the Indian Context

All these techniques for identifying Emotion are for Western music and are based on Thayer’s model, or a variation of it. In the Indian context, Emotion is normally identified, using a set of pre-defined standards called Navarasa. In addition to Thayer’s model, we have considered classification of Emotion, using this model of Navarasa including love, comedy, happiness, surprise, shyness, peace, pity, fear, anger, as shown in Figure 6.6.

In the Indian context, the term “Nava” means nine, and hence, we classify Emotion into nine classes (Figure 6.6). We have modified Thayer’s model, and created a cluster to incorporate the nine classes of Emotion, based on Carnatic music specific features.

![Figure 6.6 Navarasa way of clustering Emotions](image_url)
In our work, in addition to the features of arousal and valence that are used by Thayer’s model, we have also incorporated the characteristics of Carnatic music as features for Emotion recognition. The use of additional features for music Emotion identification using pitch multiplicity and loudness has been motivated by Yang et al (2007, 2008). We propose the use of features like Sruthi, which is conveyed as a tonic, pitch intensity and pitch fluctuations, specified as the Pitch contour. These additional features are used to identify the Emotion of the musical signal. Yet another important characteristic of Carnatic music, which is Raga, could also be used for Emotion identification, since Carnatic music suggests the use of a specific Raga to convey a set of Emotions (Sambamurthy 1983). But the disadvantage of using Raga for identifying Emotion is that one Raga could convey multiple Emotions. Therefore, in our work, we use Raga to validate and disambiguate the identified Emotions in addition to the additional features as specified earlier.

We have modified the algorithm that was used by Yeh et al (2009), by incorporating novel features to suit Carnatic music and constructed an Emotion Model for Emotion identification. In addition to this model, a GMM is constructed using the articulatory features to perform a first level of Emotion identification, by assuming a music Emotion similar to that of speech.

6.3.2 EMOTION MODEL – A MULTI-LEVEL APPROACH TO EMOTION IDENTIFICATION

The proposed Emotion identification is done by a two-level algorithm. In the first level, articulatory features are used to construct a GMM to identify the Emotion. These identified Emotions are disambiguated, using the newly designed Emotion model. From the work done by Yeh et al (2009) the use of the GMM is justified to perform a first level of Emotion
identification, and hence, we have constructed the GMM model but based on the articulatory features, which is conveyed by means of the MFCC values as a first level of Emotion identification.

The GMM is constructed for the nine Emotions as specified in Tamil literature: comedy, surprise, happiness, fear, anger, pity, love, shyness and peace. This model is used in the place of Western music’s Thayer’s model. As the GMM could result in matching one Emotion to multiple Emotions, one more new model is proposed based on the characteristic features of Carnatic music. For these nine Emotions, their characteristics, in terms of the features of tonic, pitch fluctuations, intensity and CICC, are used and based on these, a model is constructed. The values are represented in the model as fuzzy values of these features, indicated as high, low, and medium. For example, a very low tonic and a low intensity, typically convey a sad Emotion. In addition to the other features, Raga is also used (also determined from the input as explained in Chapter 5) to reduce the number of Emotions, that have been identified from the previous stage. A survey of the Emotions that a typical Raga conveys are examined and this has been used as reference in addition to that available in the literature (Chordia and Rae 2008). The proposed Emotion model is shown in Figure 6.7.

From Carnatic music literature, (Sambamurthy 1983) the typical Emotions conveyed by a Raga are made available in the Emotion model.

<table>
<thead>
<tr>
<th>Emotion</th>
<th>Tonic</th>
<th>Pitch fluctuation</th>
<th>Intensity, CICC</th>
<th>Raga</th>
</tr>
</thead>
</table>

**Figure 6.7 Emotion Model**
Therefore two models are used – a first level identification using the GMM model, and further disambiguation using the Emotion model. The GMM is trained using a data set that consists of 1500 songs including all the nine Emotions.

In this work, we have constructed the GMM for each of the nine Emotions. These nine Emotions have been given alternative names specified in Table 6.1. Based on Table 6.1 the Gaussian mixture model is tagged with the names of the main and alternative Emotion names. The Emotion model is constructed with the values of intensity, pitch fluctuations, Raga and tonic using the training input. A Raga that is typically used to convey a given Emotion is determined, based on the analysis as specified in literature (Sambamurthy 1983), and using the survey performed by Chordia and Rae (2008).

### Table 6.1 Emotions and their alternative names

<table>
<thead>
<tr>
<th>Emotion name</th>
<th>Alternative Emotions</th>
</tr>
</thead>
<tbody>
<tr>
<td>Comedy</td>
<td>High energy, patriotism, joy</td>
</tr>
<tr>
<td>Fear</td>
<td>Sadness, depression, boredom</td>
</tr>
<tr>
<td>Anger</td>
<td>Disgust</td>
</tr>
<tr>
<td>Happiness</td>
<td>Excitement, pleasure</td>
</tr>
<tr>
<td>Love</td>
<td>Anxiety</td>
</tr>
<tr>
<td>Pity</td>
<td>Relaxation, sleepiness</td>
</tr>
<tr>
<td>Surprise</td>
<td>pleasure</td>
</tr>
<tr>
<td>Shy</td>
<td>Nervousness</td>
</tr>
<tr>
<td>Peace</td>
<td>Calm, pleasantness</td>
</tr>
</tbody>
</table>

During the testing phase, features are extracted from the input signal, and Emotion is identified by comparing it with the GMM. The identified set of Emotions is disambiguated using the Emotion model. The
Earthmover’s distance was used for similarity checking, as it is typically used for music processing (Kim et al 2010, Typke et al 2003, and Typke et al 2005).

6.3.3 Analysis of the Emotion Identification Algorithms

The GMM-Emotion Model algorithm for Emotion identification and disambiguation is compared with the algorithm using the GMM alone. It has been found that the GMM based algorithm identified more than one Emotion for a given input, while our algorithm disambiguates the Emotion, where more than one Emotion has been identified, the algorithm reduces it to mostly one or two Emotions as the one being conveyed in the input music. The results are tabulated in Table 6.2. From the table it is obvious that the disambiguation algorithm was able to reduce the number of Emotions identified by atleast one, when compared with the number of Emotions identified by the GMM model based algorithm.

Table 6.2 Emotion recognition comparison

<table>
<thead>
<tr>
<th>No. of Songs</th>
<th>Emotion Tested</th>
<th>No. of Emotions identified by the GMM only</th>
<th>No. of Emotions identified by the GMM and Emotion model</th>
</tr>
</thead>
<tbody>
<tr>
<td>15</td>
<td>Comedy</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>20</td>
<td>Fear</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Anger</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>18</td>
<td>Happiness</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>20</td>
<td>Love</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>12</td>
<td>Pity</td>
<td>3</td>
<td>2</td>
</tr>
<tr>
<td>14</td>
<td>Surprise</td>
<td>3</td>
<td>1</td>
</tr>
<tr>
<td>15</td>
<td>Shy</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>18</td>
<td>Peace</td>
<td>4</td>
<td>2</td>
</tr>
</tbody>
</table>
The advantage of our algorithm is that it reduced the ambiguity in the Emotion conveyed by the input song to one or two Emotions, as compared to 3 or 4 Emotions that were identified by the GMM algorithm. Even our modified algorithm was not able to completely disambiguate the Emotions because of the use of limited number of features in the Emotion model.

The GMM based algorithm and the disambiguation based algorithm are tested with a set of 1500 songs to identify the Emotions. The output of these algorithms is compared with the input, and based on their performances the results are shown in Figure 6.8.

![Emotion Identification Comparison](image)

**Figure 6.8** Comparison of the GMM vs GMM + Emotion model algorithms for Emotion identification
It is evident that our algorithm performed better compared to the GMM algorithm in most cases as shown in Figure 6.8. The GMM based algorithm is designed based on the articulatory features only, and hence, was less accurate. This concludes that the music signal cannot be thought of as a speech signal, and hence, the use of articulatory features alone will not be able to convey Emotion from music, as it is being for speech. The disambiguation algorithm was able to reduce the number of identified Emotions, mostly based on Raga, tonic and pitch fluctuations. For example, the happiness songs are typically conveyed by a medium tonic, and would cater to the Ragas, Kalyani, Kaapi, Sree and Desh. Similarly, the Emotion of fear or sorrow Emotion is conveyed by the Raga Panduvarali, Mughari, Subapandhuvarali, while that of a romantic mood is conveyed by Kalyani, and Mohanam.

According to Carnatic music literature, one Raga can be used to convey multiple Emotions by varying the tempo, pitch fluctuations, tonic and intensity of the song. So, when the Emotions that were identified by the GMM matched with more than one Raga which is available against multiple Emotions in the Emotion model, then other features like pitch fluctuations, CICC and intensity are used to disambiguate the number of Emotions identified. The disambiguation percentage is calculated as the ratio of the number of Emotions identified by the GMM as against the combination of the GMM with Emotion model which is shown in Figure 6.9.
The error rates of both the algorithms are tabulated in Table 6.3.

Table 6.3 Comparison of emotion identification algorithms using error rate

<table>
<thead>
<tr>
<th>Emotion</th>
<th>By GMM only</th>
<th>By GMM followed by the Emotion model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Error Rate</td>
<td>22%</td>
<td>14%</td>
</tr>
</tbody>
</table>

The error rate could be further reduced by using additional features in the Emotion model, and incorporating other classifiers for Emotion identification.
6.4 GENRE IDENTIFICATION

Genre refers to the style of music. In Western music some typical Genres include Rock, Jazz, Beat, Blues, Rap, etc. In the Indian music context, many such Genres of music exist, like Thalattu, Oppari, Vayalora pattu, Ghana, Therukoothu, Melody, Romance, Keli, etc. In this thesis we propose to identify the typical Genres of music that are specific to the South Indian context.

As already discussed, the typical Genres of Tamil music, namely, Ghana (music that has a simple repeated melody which is typically used in places to necessitate ease of singing), Keli (the song symbolizes teasing), Themmangu (referring to the king of folk music), Thalattu (music used for putting some one to sleep), Vayalora pattu (music that is rendered during farming to ease out the tiredness), melody (music indicating feeling of comfort), oppari (music used to mourn the death of someone to convey sorrow), and therukoothu (typical road side music).

Based on the discussion on Chapter 2 on Genre identification and the survey carried out by Scaringella et al (2006), in this work, we wanted to explore the capability of the CICC features for Genre identification. From the conclusion that can be arrived at based on the work on Genre identification and the use of Cepstral features for the same, we wanted to explore the use of the CICC features for the Genre identification of South Indian music. The use of the CICC is justified, as it incorporates the Cepstral features representing the timbral characteristics and pitch information in the form of Tonic. Hence, we explored Genre identification of Tamil music using this CICC feature and additional information in terms of the swara pattern.
We explored two approaches, one using the simple concept of tracking the pitch envelope and the CICC, to construct the GMM to identify the Genre. The other approach used the HMM based on the CICC features for determining the Genre.

6.4.1 Genre Identification – GMM Approach

From the extracted CICC features that were determined from segments obtained from our Tala based algorithm, the mean and variance are computed using the Gaussian Mixture Model. This Gaussian Mixture Model and the pattern of the identified swaras are used to cluster and classify every Genre, which is later used for identification. This swara pattern and the GMM are used to identify the Genre of the given music signal.

6.4.2 Genre Identification – HMM Approach

In our next approach for Genre identification, from the segments of the Tala based segmentation algorithm, the CICC features are extracted. Using these features, the sequence of the CICC features are used to construct a Hidden Markov Model. To construct this Hidden Markov Model, a training corpus consisting of nearly 60 to 100 minutes of songs, belonging to each Genre is used, to observe the pattern of the CICC values. A HMM is identified using the initial probability, transition probability and output probability (Rabiner L and Juang B 1986). The initial probability is based on the probability of the first node’s CICC values, as the value corresponding to the first segment. The transition probability is established by comparing any two segments, and is the probability of these two segments occurring adjacent to each other. The output probability is also dependent on the transition probability, where we establish this value as the probability of being the last state or an intermediate state. The output probability yields the possibility of
the swara pattern that could be present as a result of this state of the HMM. Using the HMM, the Genre is identified from the input music.

### 6.4.3 Analysis of Genre Identification

The input signal is sampled, and using our algorithm as explained in Chapter 4, the CICC coefficients and the swara pattern are determined. Using these features the GMM and HMM is constructed, and is used to determine the Genre, Themmangu, Vayalora pattu, Ghana, Melody, and Romance using both the algorithms. The results are compared against the known data available from the information in the audio CD’s tag. The results are tabulated in Figure 6.10.

![Figure 6.10 Performance comparison of Genre identification](image)

From the figure it has been found that the HMM based algorithm was closer to that of the survey data available in the CD, and outperformed the GMM based algorithm. This is due to the fact that a HMM is a more
robust model compared to the single state GMM. The error rate is tabulated in Table 6.4.

**Table 6.4 Error rate of Genre identification**

<table>
<thead>
<tr>
<th>Error Rate</th>
<th>By GMM</th>
<th>By HMM</th>
</tr>
</thead>
<tbody>
<tr>
<td>24%</td>
<td></td>
<td>15%</td>
</tr>
</tbody>
</table>

The use of the CICC features is actually a representation of the timbral and pitches feature combination, and hence, the use of the CICC features increased the identification of the Genre to 76% and 85% in the GMM and HMM based approaches respectively.

After identifying the key music and non-music components from the input music signal as discussed in Chapter 5 and in this Chapter, these components can be used as key values for indexing a Music Information Retrieval system which is discussed in the next Chapter.