ISSN: 2278 – 1323
International Journal of Advanced Research in Computer Engineering & Technology (IJARCET)
Volume 5, Issue 6, June 2016

Content Based Classification of Audio Using MPEG-7 Features

ManasiChoche, Dr.SatishkumarVarma

***Abstract*—The segmentation plays important role in audio classification. The audio data can be divided into silent or non-silent segments. The non-silent segments or sub-clips are further classified into five types namely music, speech over music, pure speech, speech over environmental sound, and environmental sound. The rule based and model-based are the two types of approaches used in classification. There are various model-based methods such as Neural Network (NN), Hidden Markov Model (HMM), Support Vector Machine (SVM), and Gaussian Mixture Model (GMM) are used for classifying audio segments. In this paper, both HMM and SVM statistical models are used to classify audio segments. The MPEG-7 has four audio features namely Audio Spectrum Centroid (ASC), Audio Spectrum Spread (ASS), and Audio Spectrum Flatness (ASF), Audio Spectrum Projection (ASP).**

***Index Terms*—Audio classification, HMM, MPEG-7 features, SVM.**

I. INTRODUCTION

The audio data is used for video indexing and analysis of content. It is segmented based on the content of the audio data. The segmentation is difficult for audio data such as broadcast news. The broadcast news contains single type and mixed type audio data. The speech data and music data are example of single type classes. The speech over music and over environmental sound are example of mixed type of classes.Audio segmentation help to classify different sound class like laughing, ringing of doorbell, or dog barks. The multimedia content retrieval is performed using audio segmetation. Environmental situations can be identified using different classes of sound. It can be used in broadcast news audio, dubbing of audio.The rule based and model-based [1] are the two types of approaches used in classification. Sometimes, rule based method fails to define all rules if the given dataset is incomplete. Hence, rule based method cannot give better result for incomplete dataset and cannot produce any class. So, there are various preferred model-based approaches such as GMM, HMM, NN, and SVM. GMM does not give better accuracy for MPEG-7 features. NN model is mathematical model and more complex. HMM and SVM are referred for the better performance.

II. RELATED WORK

Many works have been done on audio classification and audio segmentation. This section gives a quick review on previous work. The SVM is used in audio segmentation [2]. Here the audio is classified into five classes. The five classes are music, silence, non-pure speech, background sound, and pure speech which include speech over music and over noise. SVM model is used to train class boundaries. A sound clip is segmented first. And then every sub-segment of 1 second is classified into above mentioned classes. The audio classification method for analysing material of film based on MPEG-7 features was carried out by Hyoung-Gook et al. [3]. This technique consists of different description scheme as low-level and high-level description. The classification of mixed type audio data using SVM is experimented in [4]. Audio features are not selected for confining properties of audio data. Instead of that various formats of every feature are designed which describes their properties. A new technique based on environmental audio data is introduced by Jia-Ching [5] Wang et al. The proposed soundclassifier is presented with the help of KNN and SVM. The audio features such as spectrum flatness, spectrum spread, and spectrum centroid are used for feature selection. The proposed classification model uses KNN and SVM methods. A model which is used for clustering makes the use of aimproved BIC (Bayesian Information Criterion) method which gives significantly high performance than KL2 (Kullback-Liebler) and it is highly efficient than BIC method. Then they used cluster models which are trained properly for labelling speaker clusters. For presenting advantages of this new algorithm

some experiments were taken [6] [8].Tin Lay et al.synthesizes spectral characteristics of audio types and presents audio power features which are based on spectral characteristics and harmonic improvement to classify audio. To enhance primitive spectral properties multi-model HMM is used. Rule based approach is used, which gives accuracy about 85.8% [7].

III. Proposed System

The architecture of content based classification of audio system is as shown in fig. 1 which consists of four phases such as Feature Extraction, Audio segmentation, Training, and Audio classification.

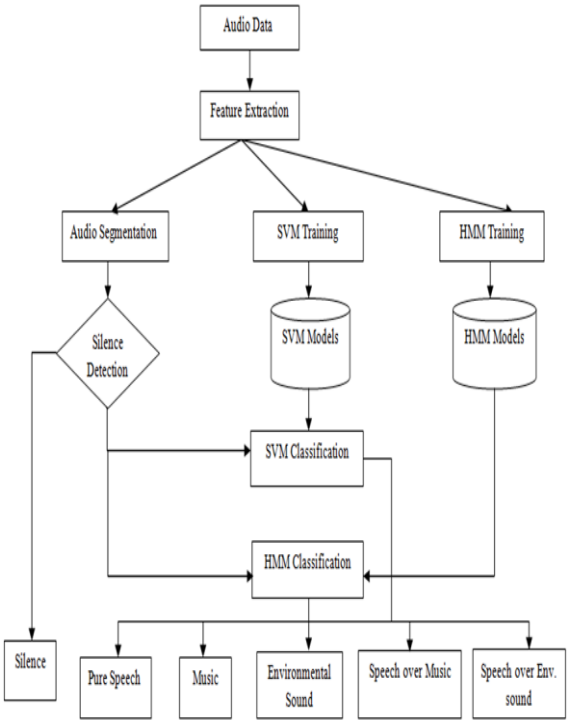


Figure 1: Architecture of content based audio classification system

A. Feature Extraction

It extracts audio clip or data into MPEG-7 Audio Power features. MPEG-7 audio power features are used in this system such as ASP, ASF, ASS, and ASC. The features are calculated based on data and frequency of an audio file.

B. Audio Segmentation

It segments an extracted audio data into silent and non-silent segments. The segmentation plays important role in segmenting audio data into homogeneous segments. Homogeneous means organizing same type of data in one segment.

C. Training

It trains different classifiers using two supervised learning algorithms SVM and HMM. The extracted features of audio data are taken into consideration for training of SVM and HMM classifiers. And these classifiers are used to identify the class of new audio data.

D. Audio Classification

It classifies trained audio data into six classes of sound. There are various preferred model-based approaches such as GMM, HMM, NN, and SVM. HMM and SVM are referred for the better performance.

This system uses two classification models:

1) Hidden Markov Model

HMM model processes with time varying characteristics. K-means algorithm for clustering is used to estimate the values. In this algorithm the centre of different clusters are initialized first. The nearer cluster to every data is determined. And then the location of every cluster is set to the mean of data points which belongs to same cluster.

2) Support Vector Machine

SVM is a classifier consists of optimal hyper-plane also known as binary classification model. In this model, the waveform of audio file is converted into individual frame segments. Every frame is then changed into features in terms of vectors. These feature vectors are given to the SVM classification model. Every frame is labelled as +1 or -1 based on decision function. And then the sum of all labels or tags is calculated. If the of audio data is greater than zero, then it is classified as +1 class. And if it is less than zero, then it is classified as -1 class.

IV. Methodology

The flowchart of proposed system is as shown in fig. 2,

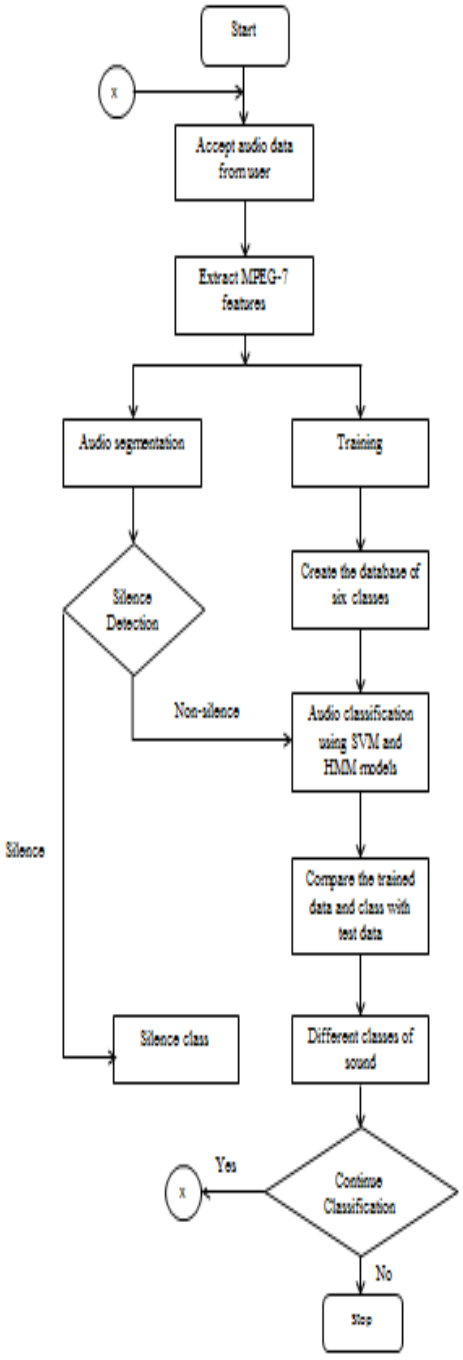


Figure 2: Flowchart of implemented algorithm

The algorithm for content based audio classification using MPEG-7 features is given below, where input will be audio sound and output will be six different classes of sound.

Input: Audio Sound

Output: Six classes of sound (music, speech over music, pure speech, speech over environmental sound, environmental sound, and silence)

Step1: Accept audio data as an input.

Step2: Extract MPEG-7 features such as ASP, ASF, ASS, and ASC.

Step3: Segment audio data into silence and non-silence by silence detection.

Step4: If it is segmented as silence, then it is considered as silence class.

Step5: Else it is segmented into one of the six types of audio.

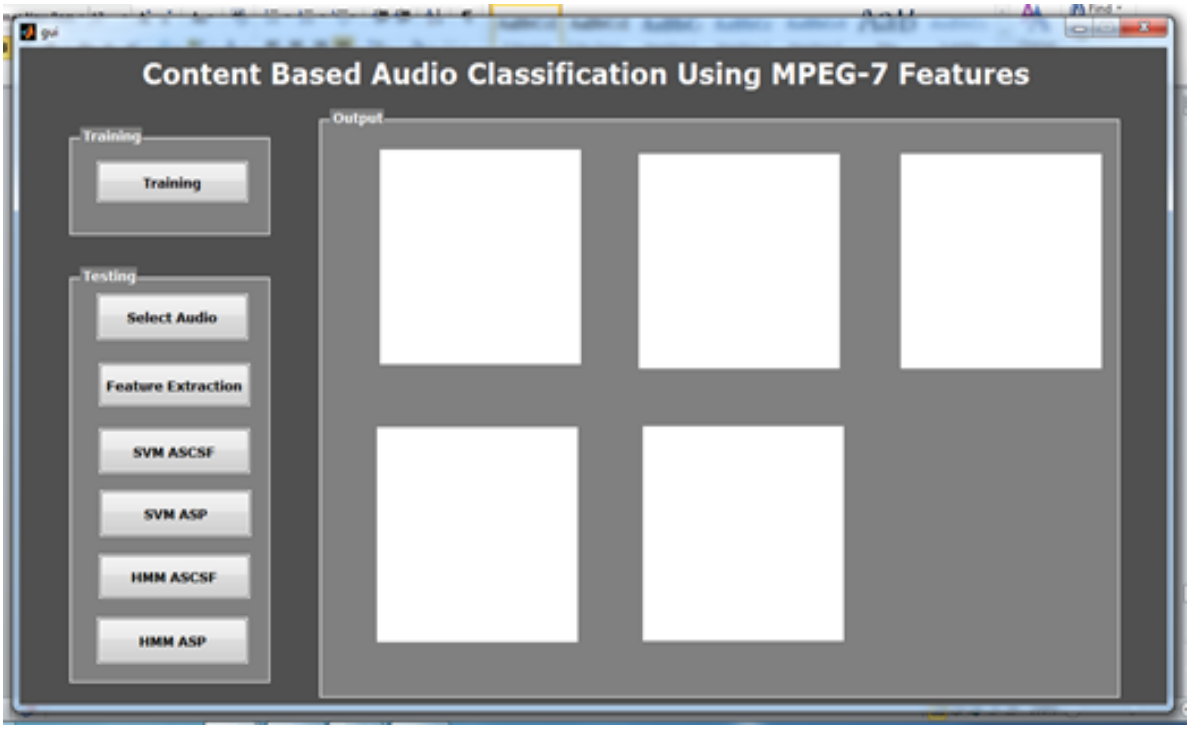
Step6: Train the classifiers by using SVM and HMM supervised learning algorithms with the help of extracted features.

Step7: Use the above trained data as input to classification module SVM and HMM.

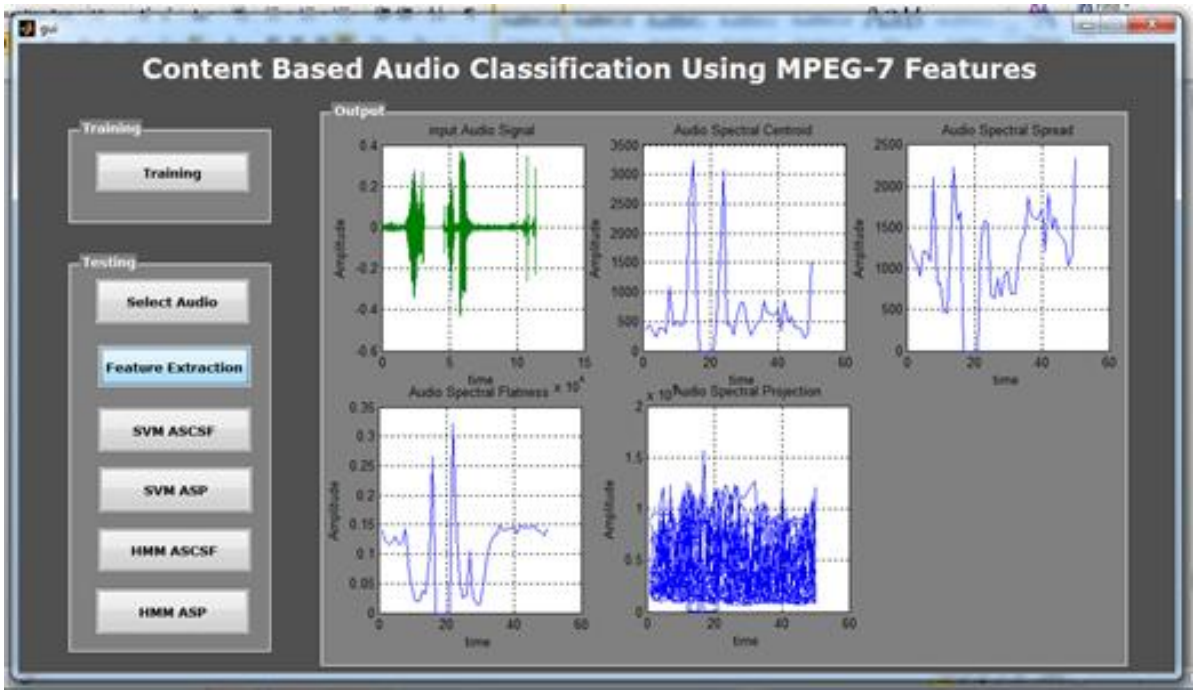
Step8: Classify audio data into six different classes as mentioned in problem statement.

V. WORKING OF SYSTEM

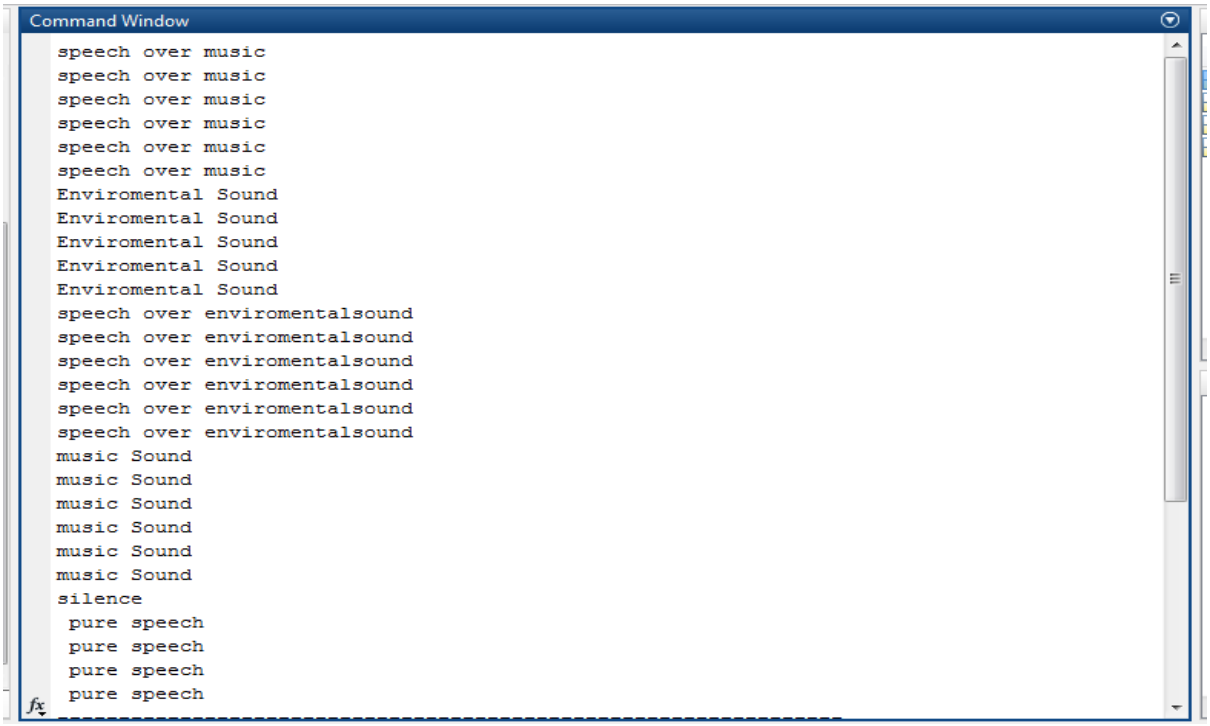
Audio files are selected as input by user. The features of these audio files are extracted. After feature extraction, the database of ASC, ASS, ASF, and ASP features is created. Then, the spectrogram is plotted on user interface screen as an output. Audio data is segmented into silence and non-silence segments. The two classifiers SVM and HMM are trained based on these features. The features of trained data and trained class are generated which is used further. The trained data, trained class, and test data are taken as input by classification module. The features of test data are compared with already created database. Finally, audio files are classified into six different classes such as pure speech, music, environmental sound, silence, speech over music, and speech over environmental sound. The working of this system is as shown in fig. 3 which consists of four parts,



(a) Main Graphical User Interface



(b) Feature Extraction



(c) SVM Classification of Audio Files

(d) HMM Classification of Audio Files

Figure 3: Working of Proposed System

VI. Experimental results

We performed experiments on different audio files and calculated rate of classification as shown in table 1,

Table 1: Classification rate of different audio classes

Name of Audio class	No. of input samples	No. of classified samples	Classification rate (%)	Average Classification rate (%)
Pure speech	10	9	90%	95%
Music	10	10	100%	
Silence	01	01	100%	
Environmental sound	10	9	90%	
Speech over music	10	10	100%	
Speech over environmental sound	10	9	90%	

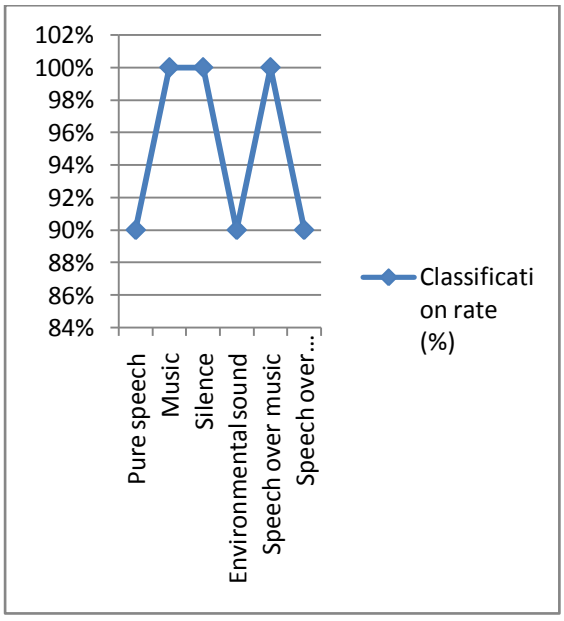


Chart 1: Classification rate of audio class

VII. Conclusion

This paper describes a system where audio is segmented into silence and non-silence segments. Then it is classified into six different classes such as music, speech over music, pure speech, and speech over environmental sound, environmental sound, and silence. Two classifiers HMM and SVM are used for classification of audio data. The HMM is preferred in MPEG-7 audio classification. The different combinations of extracted features are used to train the classification model. The performance of classification methods is analysed. The experimental results show that the average classification rate is 95%.

Acknowledgment

I thank my guide Dr.SatishkumarVarma for his guidance and suggestions during the preparation of this paper. Also I thank my colleagues and other professors for their support and cooperation which helped me to complete this study.

REFERENCES

[1] EbruDogan, Mustafa Sert and Adnan Yazici, "Content-Based Classification and Segmentation of Mixed-Type Audio by Using MPEG-7 Features,"*International conference on Advances in Multimedia*, pp. 152 – 157, 2009.

[2] Lie Lu, Stan Z. Li and Hong-Jiang Zhang, "Content-based Audio Segmentation Using Support Vector Machines", *Microsoft Research China*, pp. 749 – 752, 2003.

[3] Hyoung-Gook Kim, Nicolas Moreau and Thomas Sikora, "Audio Classification Based on MPEG-7 Spectral Basis Representations," *IEEE Transactions on Circuits and Systems for Video Technology*,Vol. 14, no. 5, pp. 716 – 725, 2004.

[4] Lei Chen, SuleGunduz and M. Tamer Ozsu, "Mixed Type Audio ClassificationWith Support Vector Machine," *IEEE Intl. Conference on Multimedia and Expo*, pp. 781 – 784, July 2006.

[5] Jia-Ching Wang, Jhing-Fa Wang, Kuok Wai He and Cheng-Shu Hsu, "Environmental Sound Classification Using Hybrid SVM/KNN Classifier and MPEG-7 Audio Low-Level Descriptor," *IEEE International Joint Conference on Neural network proceedings*, pp. 1731-1735, 2006.

[6] Hugo Meinedo andJodoNeto, "Audio Segmentation, Classification and Clustering in A Broadcast News Task", *IEEE International Conference on Acoustic, Speech and signal processing*, Vol. 2,pp. 5-8, 2003.

[7] Tin Lay and Haizhou LI, "Broadcast News Segmentation by Audio Type Analysis", *IEEE International Conference on Acoustic, Speech and signal processing*, vol. 2, pp. 1065-1068, 2005.

[8] Kalid Azad, "K-means clustering" [Online] available: http://www.onmyphd.com/?p=k-meansclustering.

ManasiChocheis pursuing M.E. in Computer Engineering from PIIT. She completed B.E. in Computer Engineering from RMCET in the year 2010. She has published papers in international journal. Her research interest is in Digital processing and Audio/video processing.
Contact no.: 9890108724

Satishkumar L. Varmahas received his B.Tech and M. Tech. degrees in June 2000 and January 2004 respectively in Computer Engineering from DBATU, Lonere, India. He has received Ph. D degree in Computer Science and Engineering from SRTMU, Nanded, India. He has worked for 2 years as Lecturer in the Department of Computer Engineering, RMCET, Devrukh, around 3 years as a Lecturer in the Department of Computer Engineering, SAKEC Mumbai and 5 years as Assistant Professor and Head of Department of Information Technology, DBIT, Mumbai. Currently he is Associate Professor and Head of Department of Information Technology, PIIT, Navi Mumbai, India. He has published one book chapter and around 20 papers in referred National as well as International Conferences and Journals including IEEE, Springer LNCS, IET, IJACSA and IJIP. His researchinterest includes Multimedia Systems, Digital Image Processing and Computer Vision.