IJRRECS/December 2013/Volume-1/Issue-8/1805-1810 **ISSN 2321-5461**

INTERNATIONAL JOURNAL OF REVIEWS ON RECENT ELECTRONICS AND COMPUTER SCIENCE

A NOVEL APPROACH FOR AUTOMATIC TEXT-INDEPENDENT SPEAKER TRACKING SYSTEM USING SOURCE FEATURE

A.Mamatha[1], V.Subba Ramaiah[2]

[1]**M.Tech Student, Dept of CSE, Mahatma Gandhi Institute of Technology, Hyderabad, A.P, India**

[2]**Senior Assistant Professor, Dept of CSE, Mahatma Gandhi Institute of Technology, Hyderabad, A.P, India**

ABSTRACT:

Speaker recognition is the procedure of automatically recognizing who is speaking based on the individual information incorporated in speech waves. The technology of speaker recognizing is used in approaches of biometric and is used in numerous commercial applications such as banking by telephone, and various security associated applications. Speaker tracking is also essential in numerous applications, such as conference and meeting indexing, retrieval of audio/video or else browsing, adaptation of speaker intended for speech recognition and video content examination. Speaker segmentation that is near to speaker tracking but there is no information concerning the individuality and number of speakers present. In the system of Text-Independent Speaker Recognition, the task of recognizing a speaker by means of machine and involves two phases: such as training and testing. The performance of identification degrades significantly due to the existence of noise and can generate a major obstruction in the design of a system of commercial recognition that is necessary to be used in standard everyday situations.

Keywords: Speaker recognition, Speaker tracking, Text-Independent, Training phase.

1. INTRODUCTION:

Speech signals are needed to be indexed based on the speaker utterance and this procedure is called system of speaker segmentation otherwise speaker indexing system. In existing systems, unidentified speaker is recognized from the specified speech signal. In numerous real time conversations and information broad casting, the speech is unremitting, opening and conclusion of speech segment of a speaker is unidentified [4]. Speaker tracking is also essential in numerous applications, such as conference and meeting indexing, retrieval of audio/video or else browsing, adaptation of speaker intended for speech recognition and video content examination. Speaker recognition is the procedure of automatically recognizing who is speaking based on the individual information incorporated in speech waves. Speaker recognition can be categorized into two tasks, specifically speaker identification and speaker verification [8]. Speaker identification is the procedure of distinguishing the speaker from the specified spoken utterance from the recorded set of N Speakers. Speaker verification is the procedure of accepting or else rejecting the individuality claim of the speaker. The technology of speaker recognizing is used in approaches of biometric and is used in numerous commercial applications such as banking by telephone, and various security associated applications [1]. Conventionally, the task of speaker recognition supposes that training and testing comprises of records of mono-speaker. Subsequently, to hold this kind of multi-speaker recordings, several extensions of the speaker recognition job are essential, such as: The detection of N-speaker which is comparable to the verification of speaker. It comprises in determining whether a set of intentional speakers in a conversation are speaking. Speaker tracking is the job of distinguishing the multiple speakers from the specified speech signal. Speaker recognition shown in fig1 is one of the centre fields to examine. Speaker segmentation that is near to speaker tracking but there is no information concerning the individuality and number of speakers present [12].

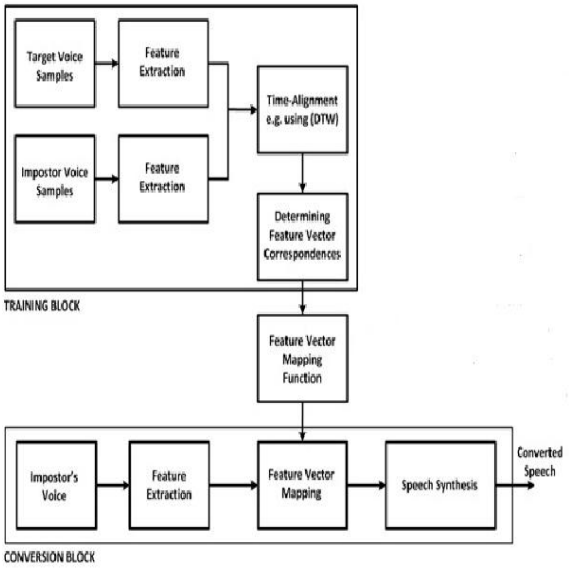

Fig1: An overview of speaker recognition.

2. METHODOLOGY:

A set of advantageous description of feature vectors intended for speaker recognition are: competent in representing the information of speaker specific; Easy to compute; steady over time; Occur regularly in speech; modify minute from one environment to another; Not vulnerable to mimicry; elevated inter-speaker difference; small intra-speaker disparity [2] [7]. In the system of Text-Independent Speaker Recognition, the task of recognizing a speaker by means of machine and involves two phases: such as training and testing. Training is a procedure of making the system is acquainted with the speakers and deals with gathering of data from the sounds of people to be recognized. Testing is the task of recognizing an unidentified utterance. This is proficient by means of making some kind of evaluation connecting the unidentified utterance and the training information. This method should work irrespective of the text moreover in process of training or testing process. The system does not contain any preceding knowledge of the text which is spoken by the person. Scheming of a system of text-independent recognition is harder than designing a system of text-dependent however has a benefit of being more flexible [10]. In the system of typical text-independent speaker recognition, each block corresponds to an exceptional constituent of the system. A system of text independent speaker recognition includes two parts such as: front-end or feature extraction as well as back-end or actual recognition [6]. These systems make use of processed appearance of speech signals as an alternative of using raw speech signal as it is gained. This is to decrease the time devoted in recognizing the speaker and to create the procedure easy by means of reducing the stream of data and exploiting its benefit of being disused. Training the representation includes ENROLL Phase which is one of significant phases used in the system of text-

independent speaker recognition engaged subsequent to the phase of feature extraction. Each model of speaker is trained by means of the extracted feature vectors and is accumulated in the trained database by means of equivalent speaker ID which is exceptional [11]. There are two types of representations that can possibly be used for training the input data. They are models of parametric and nonparametric and these models have a meticulous structure considered by a set of parameters. By defining the organization, the outline of the model has been specific and restricted to a specific constraint. This makes sure that it makes a well-organized use of the data in approximating the model constraints [3]. Another major benefit in using model of parametric is that the changes in the parameters can be effortlessly indomitable by the modifications in the data. The models of Parametric comprise Gaussian mixture models, Hidden Markov in addition to Neural Networks. The models of nonparametric are different from the parametric models like the means in which the space is dichotomized [5]. Only the negligible assumptions concerning the functions of probability density are made. Vector Quantization and Dynamic Time Warping are instances for the models of nonparametric. Vector Quantization is made used for speaker recognition of text-independent while Dynamic Time Warping is used for speaker recognition of text-dependent [9].

3. ROBUST RECOGNITION SYSTEMS:

Almost in any application of speaker recognition, the input speech signals may not forever be clean and may possibly be ruined in numerous ways. Noise may enclose unusual speech sounds, crosstalk from multiple speakers. The performance of identification degrades significantly due to the existence of noise and can generate a major obstruction in the design of a system of commercial recognition that is necessary to be used in standard everyday situations. This basis a call intended for system of robust recognition that would be able to get better recognition rates even in the existence of noisy situations or throughout the modifications in the speaker's voice appropriate to the external noise. In order to decrease the mismatch connecting test data in noisy surroundings and speech models skilled under clean circumstances, one explanation is to put in the noise practised under test circumstances to the training data.

The use of such training information contamination gives high-quality developments in a number of systems of recognition. Consequently an approach was introduced wherein the obtainable database is trained with clean and speech signals of noisy which are generated under dissimilar noise situations. The use of information contamination can in addition be supportive intended for learning algorithms to carry out improved recognition. The robust approach of our study is based on the analysis of computing speaker on comparatively short time frames of speech and this can be made used by means of any class of recognizers used and Gaussian mixture model by means of Bayes' classification rule was used for identification of speaker.

4. CONCLUSION:

Scheming of a system of text-independent recognition is harder than designing a system of text-dependent however has a benefit of being more flexible. A system of text independent speaker recognition includes two parts such as: front-end or feature extraction as well as back-end or actual recognition. There are two types of representations that can possibly be used for training the input data. Speech signals are needed to be indexed based on the speaker utterance and this procedure is called system of speaker segmentation otherwise speaker indexing system. Speaker tracking is the job of distinguishing the multiple speakers from the specified speech signal. The use of information contamination can in addition be supportive intended for learning algorithms to carry out improved recognition.

REFERENCES:

[1] Molau S., Pitz M., Schluter R., and Ney H., "Computing mel-frequency cepstral coefficients on the power spectrum", Proceedings of the IEEE International Conference on Acoustics, Speech, and Signal Processing (ICASSP), vol. 1, pp. 73-76, May 2001

[2] Roy D, and Malamud C (1997) "Speaker identification based text to audio alignment for an audioretrieval system", Proceedings of IEEE International Conference on acoustics, speech, and signal processing, pp. 1099–1102.

[3] Rajesh M. Hegde, Hema A. Murthy, and Gadde V. Ramana Rao, " Application of the modified group delay function to speaker identification and discrimination", in IEEE Trans. Acoust. Speech, Signal Processing, pp. 517-520, 2004.

[4] Mori, K. and Nakagawa, S. (2001) "Speaker change detection and speaker clustering using VQ

distortion for broadcast news speech recognition", Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, Vol. 1, Salt Lake City, USA, pp. 413-416.

[5] Seiichi Nakagawa, Kouhei Asakawa, and Longbiao Wang, "Speaker Recognition by Combining MFCC and phase information", Proceedings International Conference on Spoken Language Processing, pp. 2005-2008, 2007.

[6] L.R. Rabiner and B. H. Juang, "Fundalnentals of Speech Recognition", Prentice-Halh [13] B. L. Berg, and A. A. Beex, "Investigating Speaker Features from Very Short Speech Records", Proc. of IEEE Int'l Symposium on Circuits and Systems, ISCAS'99, Vol.3, pp.102-105,2006.

[7] Kishore Prahallad and et al., "Significance of formants from difference spectrum for speaker identification", Proceedings International Conference on Spoken Language Processing, pp. 905-908, 2006.

[8] Bonastre JF, Delacourt P, Fredouille C, Merlin T, and Wellekens C (2000) "A speaker tracking system based on speaker turn detection for NIST evaluation", Proceedings of IEEE International Conference on acoustics, speech, and signal processing, pp. 1177–1180.

[9] Chow D. and Abdulla W. H., "Robust Speaker identification Based on Perceptual Log Area Ration and Gaussian Mixture Models", Proceedings of International Conference on Spoken Language Processing, 2004.

[10] Padmanabhan M, Bahl LR, Nahamoo D, Picheny MA (1998) "Speaker clustering and transformation for speaker adaptation in speech recognition systems", IEEE Transaction on Speech Audio Process 10:19-41.

[11] Carol Y. Espy-Wilson, Sandeep M., and Srikanth V., " A New set of features for text-independent Speaker Identification", Proceedings International Conference on Spoken Language Processing, pp. 1475-1478, 2006.

[12] S. R. M. Prasanna, Cheedella S. Gupta, and B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech", Speech comm., vol. 48, pp. 1243- 1261, June 2006.