ISSN: 2321-9653

INTERNATIONAL JOURNAL FOR RESEARCH IN APPLIED SCIENCE AND ENGINEERING TECHNOLOGY (IJRASET)

Automatic Speaker Recognisation System By The Method Of Robust Formant Frequency

Abhay Kishor Tiwari[1], Niyati Shukla[2], Dr.S.K.Shrivastava[3]

[1]*Dr.C.V.RAMAN UNIVERSITY, BILASPUR, INDIA,*
MPHIL SCHOLAR

[2]*Dr.C.V.RAMAN UNIVERSITY, BILASPUR, INDIA,*
MPHIL SCHOLAR

PROFESSOR
[3]*RAJEEV GANDHI GOVT.P.G.COLLEGE, AMBIKAPUR (SURGUJA)*

Abstract: ***Speaker recognition is the process of automatically recognizing who is speaking on the basis of individual information included in speech waves. This technique makes it possible to use the speaker's voice to verify their identity and control access to services such as voice dialing, banking by telephone, telephone shopping, database access services, information services, voice mail, security control for confidential Information areas and remote access to computers. The goal of this research is to build a simple, yet complete and representative automatic speaker recognition system. Due to the limited space, we will only test our system on a mall speech database. But one can have many database files for training the system; the more files one train/teach to the system, the more accuracy is achieved. analysis of formant tracking algorithms have shown that it provides accurate formant frequency estimates for both male and female speakers for a wide range in real-time noise conditions such as multiple background speakers Robust formant tracking algorithm provides mostly smooth formant frequency estimates than RLS algorithm. The robust formant tracking algorithm recovers quickly after erroneous estimates to go back to tracking the actual formant frequencies in the speech signal, which is not the case with RLS algorithm. Because of this reason RLS algorithm shows noisy tracking. Information about the gender is not available with RLS algorithm. But the computation complexity of RLS algorithm is less as compared to robust formant tracking algorithm. There have been some problems identified with the robust formant tracker. The algorithm occasionally gives 'choppy' and oscillating formant frequency estimates. This is an undesirable result because the actual formant frequencies of speech normally vary slowly with time and have smooth transitions. This problem is only encountered when the SNR is very low and occurs due to the algorithm tracking the excess energy added outside the formant frequency regions from the background noise source.***

Keywords: ***IEEE 802.11 media-access, Data packet, Mobile Ad Hoc Environments, mobile communications. Multicasting, broadcasting.***

I.INTRODUCTION

Speaker recognition can be classified into identification and verification. Speaker/Voice identification is the process of determining which registered speaker provides a given utterance. Speaker verification, on the other hand, is the process of accepting or rejecting the identity claim of a speaker [1]. Figure 1 shows the basic structures of speaker identification and verification systems. At the highest level, all speaker recognition systems contain two main modules feature extraction and feature matching. Feature extraction is the process that extracts a small amount of data from the voice signal that can later be used

to represent each speaker. Feature matching involves the actual procedure to identify the unknown speaker by comparing extracted features from his/her voice input with the ones from a set of known speakers [2].

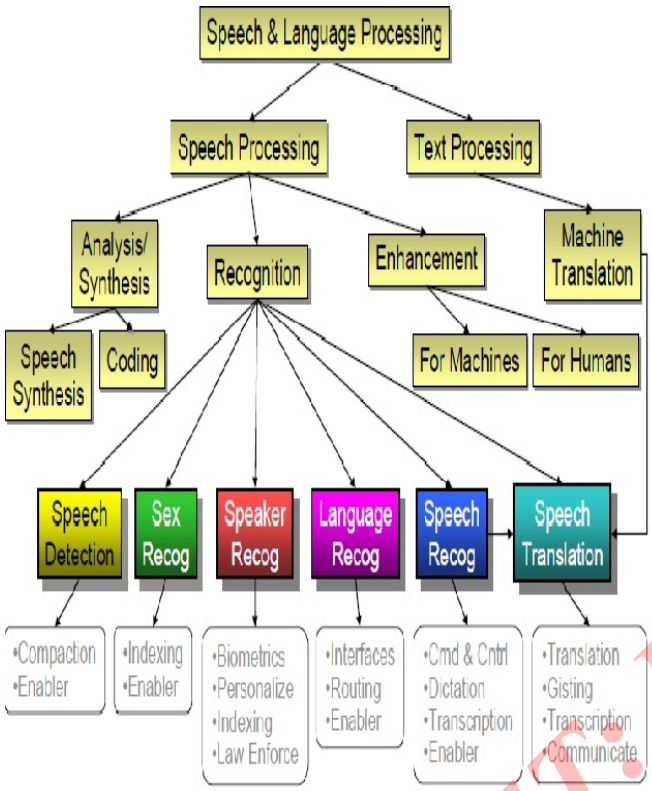

Fig1.speech and language processing (taxonomy & application)

As speaker and speech recognition system merge and speech recognition accuracy improves, the distinction between text independent and dependent applications will decrease [3]. Of the two basic tasks, text-dependent speaker verification is currently the most commercially viable and useful technology, although there has been much research conducted on both tasks. Research and development on speaker recognition methods and techniques has been undertaken for well over four decade and it continues to be an active area. Approaches have spanned from human aural and spectrogram comparisons, to simple template matching, to dynamic time-warping approaches, to more modern statistical pattern recognition approaches, such as neural networks and Hidden Markov Models (HMMs)[4]. It is interesting to note that, although striving to extract and recognize different information from the speech signal, many of the same features and techniques successfully applied to speech recognition have also been used for speaker recognition. An ideal feature would [5].

- have large between-speaker variability and small
- within-speaker variability
- be robust against noise and distortion
- occur frequently and naturally in speech
- be easy to measure from speech signal
- be difficult to impersonate/mimic
- not be affected by speaker's health or long-term variations in voice.

Speaker recognition is the task of recognizing people from their voices. Strictly speaking there is a difference between speaker recognition (recognizing who is speaking) and speech recognition (recognizing what is being said)[6]. Speaker recognition system is categorized into speaker verification (to authenticate a claimed speaker identity from a voice signal based on speaker-specific characteristics reflected in spoken words) and speaker identification (to find the identity of a talker, in a known population of talkers, using the speech input).Speaker identification is the task of determining an unknown speaker's identity[7].

II.RELETED WORK

The progress of automatic speech recognition (ASR) technology in the past 50 years can be summarized as follows

In Year 2007. The most commonly used acoustic vectors are Mel Frequency Cepstral Coefficients (MFCC), Linear Prediction Cepstral Coefficients (LPCC) and Perceptual Linear Prediction Cepstral (PLPC) Coefficients and zero crossing coefficients. All these features are based on the spectral information derived from a short time windowed segment of speech. They differ mainly in the detail of the power spectrum representation.

In Year 2008. the objective of modeling technique is to generate speaker models using speaker-specific feature vectors. Such models will have enhanced speaker-specific information at reduced data rate. This is achieved by exploiting the working principles of the modeling techniques. Earlier studies on speaker

recognition used direct template matching between training and testing data.

In Year 2009. it is common to reduce the number of training feature vectors by some modeling technique like clustering. The cluster centers are known as code vectors and the set of code vectors is known as codebook. The most well-known codebook generation algorithm is the K-means algorithm In order to model the statistical variations, the hidden Markov model (HMM) for text-dependent speaker recognition was studied.

In Year 2010. The Speaker Recognition is a major task when security applications through speech input are needed. Nevertheless, speech variability is a main degradation factor in speaker recognition tasks. Both intra-speaker and external variability sources produce mismatch between training and testing phases.

In Year 2011(Kumar *et al*.),described that the BFCC features perform well for text dependent speaker verification systems. Revised perceptual linear prediction was proposed by for the purpose of identifying the spoken language; Revised Perceptual Linear Prediction Coefficients (RPLP) was obtained from combination of MFCC and PLP.

In Year 2012 a robust way of dealing with unwanted variation. For modeling, exclusive use of inner product-based and cepstral systems produced a language-independent computationally scalable system. For robustness, systems that captured spectral and prosodic information, modeled nuisance subspaces using multiple novel methods, and fused scores of multiple systems were implemented.

In Year 2013(V. Anantha Natarajan *et al*.) addresses the issues in segmentation of continuous speech into sub-word units of speech using Formants and support vector machines (SVMs). Many studies have been conducted to identify and discriminate vowels and consonants using acoustic/articulatory differences. In this study the continuous speech is segmented into smaller speech units and each unit is classified either consonant or vowel using the Formant frequencies. This process when further combined with recognition of each unit will form a complete speech recognition system. The proposed detection strategy is tested with the speech signals recorded from the television broadcast.

In the last 50 years, research in speech and speaker recognition has been intensively carried out worldwide, spurred on by advances in signal processing, algorithms, architectures, and hardware. The technological progress in the 50 years can be summarized by the following changes [8,9]:
(1) from template matching to corpus-base statistical modeling,HMM and n-grams,
(2) from filter bank/spectral resonance to cepstral features
(3) from heuristic time-normalization to DTW/DP matching,
(4) from "distance"-based to likelihood-based methods,
(5) from maximum likelihood to discriminative approach, e.g. MCE/GPD and MMI,
(6) from isolated word to continuous speech recognition,
(7) from small vocabulary to large vocabulary recognition,
(8) from context-independent units to context-dependent units for recognition,
(9) from clean speech to noisy/telephone speech recognition,
(10) from single speaker to speaker-independent/adaptive recognition,
(11) from monologue to dialogue/conversation recognition,
(12) from read speech to spontaneous speech recognition,
(13) from recognition to understanding,
(14) from single-modality (audio signal only) to multimodal (audio/visual) speech recognition,
(15) from hardware recognizer to software recognizer, and
(16) from no commercial application to many practical commercial applications.

III.PURPOSE of RESEARCH

(1) speech systems avoid an explicit interpretation of the spectral envelope in terms of formant.
(2) The exact causes of the many-to many mapping between spectral maxima and true formants need not concern us here. What is essential is that despite numerous attempts to build accurate and reliable automatic formant extractors.
(3) The voice source may also contain spectral peaks and valleys that may affect the spectral peaks in the corresponding speech signals.
(4) it is safe to assume that if a formant-like representation fails to approach the same vowel classification performance as the true formants, it is highly unlikely that such a

representation could yield the theoretical advantage expected from true formants on a more realistic continuous speech recognition task[10].

IV. EXPERIMENT and IMPLEMENTATION. ROBUST FORMANT TRACKING ALGORITHM

The robust formant tracking algorithm discussed in the present work is the most accurate formant tracking algorithm. This algorithm is robust and accurate in continuous speech and mitigates the effects of speaker variability and different background noises. This allows the algorithm to operate independently and provide reliable formant frequency estimates for contrast enhanced frequency shaping (CEFS) amplification and other applications. shows a block diagram of the Robust Formant Tracker[11].

The speech signal is first pre-emphasized using a high-pass filter to equalize the energy and remove the spectral tilt of the speech signal. An approximate, analytic version of the signal is then calculated to increase spectral accuracy for the formant estimates through an approximate Hilbert transformer. The analytic signal is then filtered into four different bands using a bank of adaptive band-pass filters (called Formant Filters). Each of the four formant filters (F1, F2, F3 and F4) in the filter bank is made up of an All- Zero Filter (AZF) and a Dynamic Tracking Filter (DTF) [12]. The zeros of each of the AZF's are set to the latest estimate of the formant frequencies from the other three bands. The DTF provides the single pole located at the latest estimate of the formant frequency for that band. This cascade arrangement results in each of the filters having a pole around its own formant frequency and zeros at the other formant frequency locations.

Each of the four band-pass filters allows only the signal around the frequency region of the desired formant to pass through and suppresses the other frequency regions. The formant filter bank has a fundamental modification that the F1 filter of the filter bank has an added zero at the pitch frequency (F0) for further suppression of the region below the F1 frequency (the pitch region). This decreases the effects of the pitch on the F1 estimate [14].

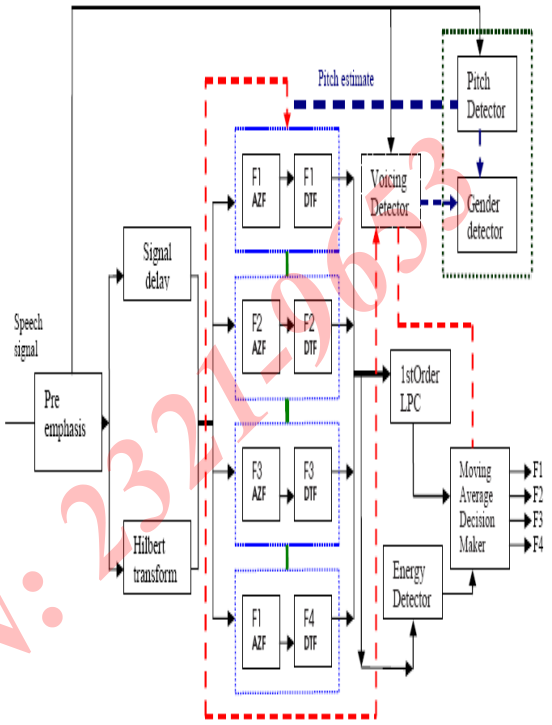

Fig.2: Robust Formant Tracker

Extensive testing of the robust formant tracking algorithm has been done which showed that the formant tracking algorithm is robust to a wide variety of real-time background noise conditions. The algorithm is able to provide reliable formant frequency estimates from continuous speech for both male and female speakers. It recovers quickly and with minimal error when problems do occur and when there is a switch in speakers [15]

APPLICATIONS OF FORMANT TRACKING ALGORITHMS

Formant frequencies play a major role in vowel identification and are also important for consonant identification. Formant tracking algorithms estimates the formant frequencies accurately. Accurate formant frequency estimates can be used for a variety of applications.

(a)These algorithms provides contrast enhanced frequency shaping amplification for hearing aids.

ISSN: 2321-9653

(b)Formant frequencies have been used to make natural sounding computer synthesized speech.
(c)Formant frequency estimates can be used for speech recognition.
(d)Formant estimates can be used in speech coding.
(e)The formant tracking algorithm can also be used for concatenation synthesis of speech

V.PERFORMANCE EVALUATIONS AND RESULTS

It has been observed that in real-life, there is often more than just one speaker present in an environment. The algorithm was also tested for the environment in which background speaker is present by estimating formant frequencies for the dominant speaker in the presence the background speakers. Different cases are considered here[16]. Testing is done with the female background speaker, multiple background speakers. Here the background speaker serves as the 'noise source'. This will cause the algorithm to start tracking the formant frequencies of the background speaker instead of those of the primary (more dominant) speaker.

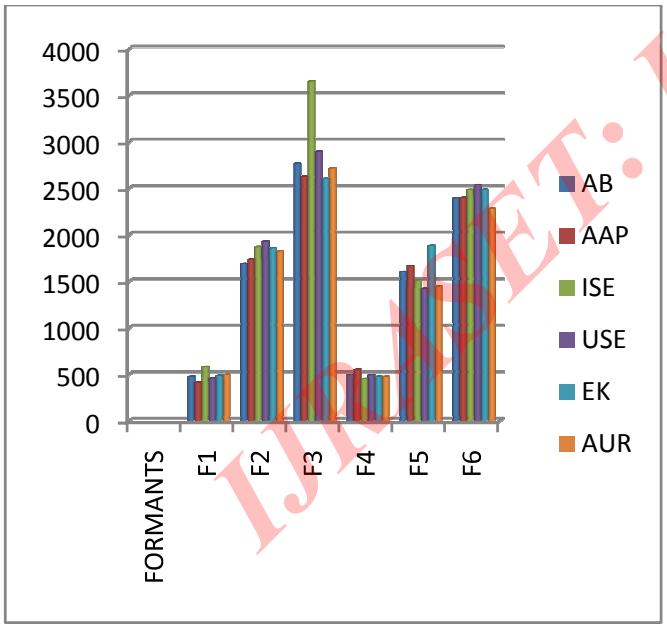

Fig.3: The loudness of the background speaker often varies in real-life

VI.APPLICATIONS OF MOBILE AD HOC NETWORK

The applications of speaker recognition technology are quite varied and continually growing. Below is an outline of some broad areas where speaker recognition technology has been or is

currently used (this is not an exhaustive list). A search on the web will produce numerous pointers to companies and products (some lists can be found at,

Access Control: Originally for physical facilities, more recent applications are for controlling access to computer networks (add biometric factor to usual password and/or token) or websites (thwart password sharing for access to subscription sites). Also used for automated password reset services.
Transaction Authentication: For telephone banking, in addition to account access control, higher levels of verification can be used for more sensitive transactions. More recent applications are in user verification for remote electronic and mobile purchases (e- and m-commerce).
Law Enforcement: Some applications are home-parole monitoring (call parolees at random times to verify they are at home) and prison call monitoring (validate inmate prior to outbound call). There has also been discussion of using automatic systems to corroborate aural/spectral inspections of voice samples for forensic analysis.
Speech Data Management: In voice mail browsing or intelligent answering machines, use speaker recognition to label incoming voice mail with speaker name for browsing and/or action (personal reply). For speech skimming or audio mining applications, annotate recorded meetings or video with speaker labels for quick indexing and filing.
Personalization: In voice-web or device customization, store and retrieve personal setting/preferences based on user verification for multi-user site or device (car climate and radio settings). There is also interest in using recognition techniques for directed advertisement or services, where, for example, repeat users could be recognized or advertisements focused based on recognition of broad speaker characteristics (e.g. gender or age).

VII.CONCLUSION

The advantages of The algorithms discussed in the present work are geared primarily towards use for CEFS amplification. It was identified earlier that in order to apply CEFS to continuous speech the second formant frequency has to be estimated accurately and in real-time. Furthermore, the estimated formant frequencies have to be smooth and the algorithm has to be able to identify formant transitions accurately so that the proper frequency dependent amplification is applied to the speech signal. Testing on the algorithms has shown that the formant frequency estimates are smooth and the formant frequency transitions are tracked accurately. Therefore, the formant tracking algorithms presented here can be used to implement CEFS amplification. With RLS algorithm, the formant frequency estimates are not very smooth and also have large jumps. Because of this reason robust formant tracking algorithm is a better choice for CEFS amplification.

VIII. FUTURE WORK

In future scope, improvement may be to modify the formant pre-filters to have variable bandwidths that are dependent on the magnitudes of the poles estimated by the linear prediction coefficients. This may further improve the formant estimates during rapid formant transitions at high SNR's, but the performance at low SNR's would likely remain unchanged.

REFERENCES

[1] K. Mustafa and I. C. Bruce, 2004 *Robust* formant tracking for continuous speech with speaker variability, in Proceedings of the Seventh International Symposium on Signal Processing and Its Applications (ISSPA), Vol. 2. Piscataway, NJ: IEEE, ,

[2] L.R.Rabiner and B.H Juang. 1993 *Fundamentals of Speech Recognition* Prentice-Hall, Englewood Cliffs, NJ.

[3] M.G. Sumithra, 2K. Thanuskodi and 3A. Helen Jenifer Archana 2005 *A New Speaker Recognition System with Combined Feature Extraction Techniques* Journal of Computer Science 7 (4): 459-465, 2011ISSN 1549-3636 Science Publications

[4] Aronowitz, H., Burshtein, D., Amir, A., 2004. *Speaker indexing in audio archives using test utterance gaussian mixture modeling*. In: Proc. Of ICSLP,

[5] Liu, E. Shriberg, A. Stolcke, D. Hillard, M. Ostendorf,and M. Harper. 2006. *Enriching speech recognition with automatic detection of sentence boundaries and disfluencies*. IEEE Trans. Audio, Speech and Language Processing, 14(5):1526–1540.

[6] I. Mporas and T. Ganchev, 2007Estimation of unknown speakers height from speech," International Journal of Speech Technology, vol. 12, no. 4

[7] Y. Liu, et. al.2005, *Structural metadata research in the EARS program*, Proc. ICASSP,

[8] Kuldeep Kumar and R.K. Aggarwal, 2010 *Hındı speech recognıtıon system usıng HTK*, International Journal of Computing and Business Research, vol.2, no.2, 2010

[9] Satya Dharanipragada, et.al206, *Gaussian mixture models with covariance s or Precisions inshared multiple subspaces*, IEEE Transactions on Audio, Speech and Language Processing,vol.14, no.4

[10]Mathias De-Wachter, et.al.,2007 *Template based continuous speech recognition*, IEEE transactions on Audio, speech and Language processing, vol.15,no.4, .

[11] Yifan Gong, 1997 *Stochastic Trajectory Modeling and Sentence Searching for continuousSpeech Recognition*, IEEE Transactions On Speech And Audio Processing, vol.5,no.1,.

[12] George Saon and Mukund Padmanabhan, 2001 *Data-Driven Approach to Designing Compound Words for continuous Speech Recognition*, IEEE Transactions On Speech And AudioProcessing, vol. 9, no.4,.

[13] Kevin M.Indrebo, et.al, 2008 *Minimum mean squared error estimation of mel-frequency cepstral co-efficients using a Novel Distortion model*, IEEE Transactions On Audio,Speech AndLanguage Processing, vol.16, no.1,.

[14] Xiong Xiao, 2008*Normalisaton of the speech modulation spectra for robust speech recognition*,IEEE transactions on Audio, Speech and Language Processing, vol.16, no.1,

[15] Mohit Dua, R.K.Aggarwal, Virender Kadyan and Shelza Dua, 2012. *Punjabi Automatic SpeechRecognition Using HTK*, International Journal of Computer Science Issues, vol.9, no.4,

[16] Sadaoki Furui, 2005 50 *years of Progress in speech and Speaker Recognition Research*, ECTI Transactions on Computer and Information Technology,vol.1, no.2,