International Research Journal of Engineering and Technology (IRJET) e-ISSN: 2395 -0056
Volume: 02 Issue: 03 | June-2015 www.irjet.net p-ISSN: 2395-0072

A New Speaker Recognition System with Combined Feature Extraction Techniques in Continuous Speech

[1]Pooja Jaiswal, [2] Praveen Chouksey,[3]Rohit Miri

MTech Scholar, Assistant Professor, HOD

[1, 2, 3] Department of CSE, Dr.C.V. Raman University(C.G.)

Abstract— words information can be conked out behind into speech mix, professor finding and words discovery. mix refers to the process of repeatedly create speech using a computer. The other main areas of see both involve language as enter: whereas the objective of professor discovery is to make out an entity based on his or her voice, words discovery effort to regularly know the linguistic satisfied of such an word. Human easily and efficiently correspond in order via language despite a lot of difficulty, including surroundings noise, slip related to natural language(stumble, full pause, false starts, etc.) and the inborn contradiction of person words. The latter stand for the most confront for custom language discovery, and will be investigative from the viewpoint of the following three areas of linguistic learn:

1. Phonetics: look and Acoustics
2. Phonology: Phonemes, Phonotactics and Coarticulation
3. Prosody: worry, pasture and beat

A professor verification method is collected of two unlike phase, a education stage and a test phase. Each of them can be see as a sequence of independent module. The first and leading unit is the constituent removal unit assigning professor in order takes out from the words. This is the plinth unit, where the entire scheme appearance relies. The next unit is lecturer model unit; represent that speaker's say and audio skin. The compilation of model is mostly needy on the type of words to be used, desired arrangement, the ease of education and updating and storage space and calculation consideration. The final module is for formation choice based on the education and testing stage. The system, in turn, production a dual choice: Either admit or reject the strength for the claim professor. conquest in speaker guarantee depends on extract and model the professor needy individuality of the language signal which can professionally distinguish one talker from another.

KEYWORDS: - Speech skill, stream of sound, desired performance, efficient message, graphical user interface

I. INTRODUCTION

In general, the aim of the lecturer credit task is to decide whether two utterance come from the same lecturer or not. present speaker detection benchmarks say a series of true and fake trials in which one has to calculate a score expressing the degree of comparison between each pair of training and testing utterances. The performance of a system can be predictable computing the equal error rate (EER) associated with false alarm and false rejects errors caused by this system Language

Feature removal is the most significant step in any routine speaker recognition system. In the last 60 years a lot of study has gone into parametric depiction of these speech skin. Several techniques have evolve one after the other in order to defeat the shortcoming of the previous one. Yet routine Speaker detection still remains a confront mainly due to variations in speaker's oral tract with time and fitness, varying ecological conditions, variation in the actions and quality of speech recorders etc. Although Mel Frequency Cepstral Coefficients (MFCC) has become a standard for speaker credit, the conservative MFCC has a poor credit in presence of sound. In this paper MFCC method was used for routine lecturer recognition in case of a slightly noisy surroundings. In this experiment a VQ codebook was shaped by clustering the education features of 9 speakers. This data was stored in a speaker folder. Here the K means algorithm was used for cluster. A twist measure based on the least Euclidean reserve was used for speaker detection.

1.1 Phonemes

Although the acoustic property distinguishing language sounds are often very slight, efficient letter is still possible because each words relies on a small subset of these sound [3,4]. Only the sounds in this subset, known as phonemes, actually serve to difference meaning within the given words.

1.2 Phonotactics

language is a continuous brook of sound without partition such as those conservatively indicated by spaces on a on paper page. A lecturer of a fussy language intuitively knows which sounds can follow which, and mechanically divides speech up into its ingredient words. listen to foreign speech,

on the other hand, is an completely different experience because these constraint, known as phonotactic rules, are not known for the words: it just sounds like a stream of nonsense [5].The partition of continuous speech into words is called segmentation, and, while phonotactics can help, it is one more source of complexity in speech recognition [6].

1.3 Co expression and connected phenomenon

To further make difficult the state of affairs, the sound comprise language are not verbal separately of one another. If this were not the case, the competence of human message would be gravely compromise because each phoneme would have to be carefully pronounced in separation, involving significantly more articulator effort. fast speech is able of thirty phonemes per second, but these are not independently perceived. In fact, the person brain is only capable of processing roughly half this figure of distinct sounds in a second [7,8].This limits effect on competence is minimized by overlap sounds while speaking, and human do so regularly. This occurrence, known as co articulation, considerably influences the acoustics of language.

1.4 Prosody: The Music of words

Prosody refers to those skin of speech which are dispersed over more than a single phoneme. It encompasses stress, pitch (tone and intonation) and tempo, and considerably contributes to the inconsistency of speech [8].

1.5 Stress

Stress refers to the articulation of a syllable with moderately high or low loudness. In Hindi, stress now and then serves to differentiate between functions of a word. Acoustically language, stress is related to the original frequency, intensity and period of the sound wave corresponding to the syllable in question [9].

1.6 Pitch

The fundamental frequency is apparent as the pitch of a speaker voice. By varying the stress and position of the verbal folds, the field can be adjusted for a single syllable (tone) or incessantly over an whole sentence (intonation) [10].

1.7 Tempo

Tempo refers to the period of a speech sound. This timing sometimes conveys non-linguistic in order, but also serves to differentiate between word meanings in certain languages [11].

II. RELATED WORK

In Year 2008 (Liu *et al.*).Planned that the object of model method is to make lecturer models using speaker-specific quality vectors. Such model will have better speaker-specific in arrange at précis data rate. This is achieve by develop the working values of the replica technique. Earlier study on professor discovery used direct outline identical between teaching and testing information In the nonstop pattern matching, training and trying characteristic vectors are directly compare using similarity measure.

(Krause *et al*) describe that in HMM, time-dependent parameter are examination signs. study signs are shaped by VQ codebook label. Nonstop possibility events are shaped using Gaussian mixture models (GMMs). The main belief of HMM is that the present state depends on the before state.

In Year 2009(Mporas *et al*., Ming *et al*., 2007)intended that the disadvantage of pattern similar is time overwhelming, as the number of feature vectors add to. For this cause, it is general to reduce the number of education quality vectors by some modeling method like cluster. The cluster centre are recognized as code vectors and the set of code vectors is known as codebook. The most well-known codebook formation algorithm is the K-means algorithm In sort to replica the numerical difference, the hidden Markov model (HMM) for text-dependent lecturer detection was studied [13].

In Year 2010(J. Ortega-García *et al*) **describe** that professor appreciation is a main task when defense application through words input are needed. however, language variability is a main deprivation factor in professor appreciation tasks. Both intra-speaker and outside inconsistency sources make dissimilarity between training and testing phase[30].

In Year 2011(Kumar *et al*.),describe that the BFCC feature do fine for way needy professor certification systems. revise perceptual linear predict was planned by for the reason of recognize the oral speech; revise Perceptual Linear forecast Coefficients was obtain from mixture of MFCC and PLP[12].

In Year 2012(Douglas Sturim *et al.*)describe that in the professor discovery society has sustained lecture to method of justifying variation irritation. cellular phone and auxiliary-microphone record speech emphasize the need for a fit way

of selling with unwanted divergence. The plan of recent2010 NIST-SRE lecturer detection assessment reflects this examine accent. In this the tasks of the 2010 NIST-SRE with two major goal—language-independent scalable model and healthy irritation easing. For replica limited use of internal product-based and campestral systems formed a language-independent computationally scalable structure. For toughness, systems that imprison ethereal and prosodic in turn, model annoyance subspaces using many novel method, and fused scores of many systems were execute [17].

In Year 2013(V. Anantha Natarajan *et al*.) address the issue in segmen-tation of wildly words into sub-word units of words using Formants and sustain vector equipment. Many study have been conduct to be familiar with and distinguish vowels and consonants using audio/articulator dissimilarity. In this learn the nonstop speech is segmented into lesser words units and each unit is confidential either consonant or vowel using the Formant frequencies. This procedure when further joint with discovery of each unit will form a total language detection system. The proposed finding plan is tested with the chatting signals evidence from the small screen show.

The most usually used audio vectors are Mel incidence Cepstral Coefficients , Linear predict Cepstral Coefficients and Perceptual Linear forecast Campestral Coefficients and zero crossing coefficients (Yegnanarayana *et al*., 2005; Vogt *et al*., 2005). All these skin tone are based on the troubled in order derived from a short time windowed section of words. They be different mostly in the feature of the power range sign. A new alteration of Mel-Frequency Cepstral Coefficient (MFCC) characteristic has been planned for removal of speech features for professor verification (LV) application (Saha and Yadhunandan, 2000).This is evaluate with unique MFCC based feature extraction method and also on one of the new alter. The learn uses multi-dimensional F -ratio as appearance determine in lecturer gratitude (LR) application to evaluate discriminative ability of unlike multi limitation method. An MFCC like feature based on the Bark level is shown to give up alike routine in language appreciation experiment as MFCC (Aronowitz *et al*., 2005).

In Year 2014. Generate professor model using speaker-specific mark vectors. Such replica will have better speaker-specific in order at précis data rate. This is achieve by use the working values of the model method. previous study on professor detection used through pattern similar between education and difficult data. In the straight pattern similar, teaching and difficult feature vectors are straight compared using similarity assess. For the similar compute, any of the method like ghostly or Euclidean reserve or Mahalanobis distance is used (Liu *et al*., 2006).

The disadvantage of prototype alike is that it is time overriding, as the number of trait vectors increase. For this reason, it is general to decrease the figure of education excellence vectors by some model do like cluster. The collection centre is well-known as code vectors and the set of scheme vectors is well-known as codebook. The majority well-known codebook invention algorithm is the K-means algorithm (Mporas *et al*., 2007; Ming *et al*., 2007). In 1985, Soong *et al*. use the LBG algorithm for make speaker-based vector quanti zation (VQ) codebooks for professor discovery. In place to copy the arithmetical difference, the unseen Markov model (HMM) for text-dependent lecturer detection was intended. The scheme act in neural network based network were also calculated (Clarkson *et al*., 2006). In HMM, time-dependented parameter are check signs. inspection signs are shaped by VQ codebook label. constant chance trial are create using Gaussian mixtures models (GMMs) (Krause and Gazit, 2006) . The main statement of HMM is that the topical state depend on the before state.

In 1995, Reynolds proposed Gaussian mixture modeling (GMM) classifier for lecturer detection task (Krause and Gazit, 2006; Clarkson *et al*., 2006). This is the the mass broadly used probabilistic method in lecturer finding. The GMM wants enough information to replica the lecturer and hence fine production. In the GMM model system, the allocation of value vectors is model by the parameters mean, covariance and load.

III.PROBLEM IDENTIFICATION

A New professor detection System with joint feature removal Techniques in constant Speech.The problem meaning says that to non parametric representation of language signal are mostly usually used. Still if the guess of the troubled cover is resulting from a parametric estimator such as Linear analytical code LPC which can be linked to the source-filter copy of audio speech produce, speech systems avoid an open explanation of the ghostly envelope in terms of formant.

IV. PROPOSED APPROACH

Language detection technique- The plan of language discovery is for a device to be able to "listen to," identify," and "perform " oral in turn. The initial words discovery systems were first attempt in the early 1950s at Bell Laboratories, Davis, Biddulph and Balashek developed a

cut off number detection scheme for a only lecturer [14]. The object of routine lecturer detections to examine, remove set apart and recognize in order about the professor identity. The lecturer discovery scheme may be view as operational in a four stages

1. study
2. characteristic removal
3. model
4. trying

1. language study method

language in order enclose unlike type of in order that show a lecturer uniqueness. This include professor exact in sequence due to oral tract, excitation basis and recital class. The in order about the performance characteristic also fixed in sign and that can be worn for lecturer detection. The language study stage deals with stage with appropriate limit size for segmenting language signal for further revise and extracting [15]. The language analysis method done with following three technique

(a) Segmentation analysis
(b) Sub segmental analysis
(c) Supra segmental analysis
(d) Performance of System

2. Characteristic removal method

The words quality removal in a cataloging intricacy is about falling the dimensionality of the input vector while keep the discerning control of the sign [16]. As we identify from prime configuration of lecturer recognition and confirmation scheme, that the number of training and test vector wanted for the categorization complexity grow with the measurement of the given input so we need characteristic removal of words signal. Following are some feature removal.

Referring to the diagram above, you can see that the enter words will pass through two main stage in sort to get the professor quality, they are:

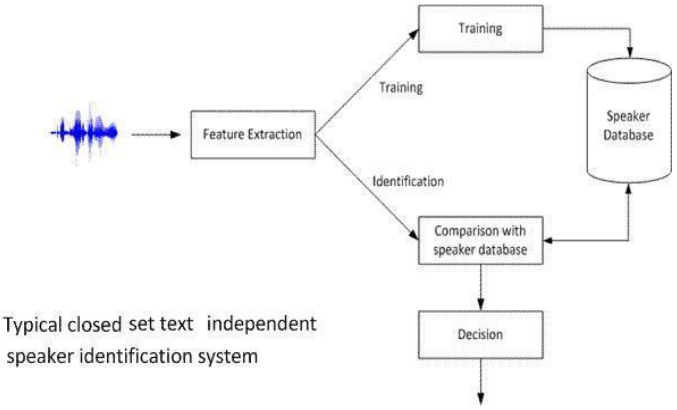

Fig 1 : symbol chart of the closed-set lecturer recognition system.

1- Characteristic elimination.
2- Category and feature similar.

Formant occurrence technique - Formants are the booming frequencies of the spoken area when vowels are exact. Formants can be recognized where there are big attentiveness or peak of power in the spectrogram analysis of a verbal example. In position to put into practice Contrast Enhanced Frequency Shaping addition in hearing aids for wildly speech, the second formant occurrence (F2) wants to be precisely predictable for verbal language [1,17]. precise formant appraisal for nonstop speech (in real time sound environment)is a tackle because formant frequencies are not easy to path in such a lively surrounds. The formant opinion algorithm wants to be well and be able to function in a wide variety of real-time sound scenarios. It must also be able to get well swiftly if it encounter any intricacy and after period of silence. For guess formant frequencies two algorithms have been available. One is healthy formant path algorithm and one more is using RLS algorithm.

ROBUST FORMANT TRACKING ALGORITHM – The robust formant tracking algorithm talk about in the there work is the mainly precise formant tracking algorithm. This algorithm is robust and precise in nonstop speech and mitigates the belongings of lecturer unpredictability and different backdrop sound. This allows the algorithm to function separately and give dependable formant occurrence approximation for dissimilarity improved frequency shaping extension and other application. Figure.2 show a block diagram of the Robust Formant Tracker [18].

The words sign is first pre-emphasized by a high-pass sieve to make equivalent the power and take away the ghostly tilt of the speech signal. An estimated, logical account of the sign

is then future to add to spectral rightness for the formant approximation through an estimated Hilbert transformer [19].

The logical signal is then clean into four unlike band by a bank of adaptive band-pass sieve (called Formant Filters). Each of the four formant filter (F1, F2, F3 and F4) in the filter bank is total up of an All- Zero Filter (AZF) and a active track Filter . The zeros of each of the AZF's are place to the newest estimate of the formant frequencies as of the other three bands. The DTF give the single pole situated at the latest approximation of the formant incidence for that band. This flow agreement consequences in each of the filters having a edge around its own formant occurrence and zeros at the other formant incidence position.

Each of the four band-pass filter allow only the mark around the incidence area of the chosen formant to pass through and suppress the other rate region. The formant clean bank has a basic modification that the F1 filter of the filter bank has an extra zero at the pitch rate (F0) for more control of the area below the F1 incidence (the pitch region). This reduce the belongings of the pitch on the F1 guess.

A first-order Linear forecast Coefficient is then intended for the logical gesture in each of the four band. From each of these coefficients a formant pre-filters are modified to track them by altering their pole and zero location. Due to the band-pass pre-filtering of each formant incidence region past to LPC, the occurrence estimation provide by LPC are more exact and the algorithm is less vulnerable to errors due to setting sound.

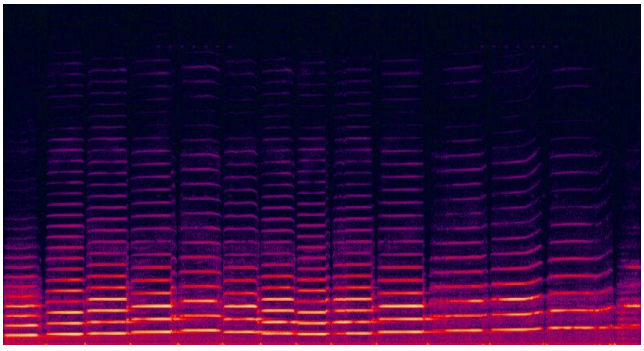

Fig 2. Spectrogram of the value for man.

V RESULT

The main object of formant tracking algorithms is to get bigger a dependable formant track algorithm that is healthy in concurrent sound scenario. Dissimilar test cases are explain and the presentation of the algorithms under these condition has been discuss. The formant tracking algorithms has been knowledgeable using unsynthesized language signal. Testing by unsynthesized verdict allow quantitative analysis of the appearance of the formant tracker because the formant frequency value of the unsynthesized words signals are not known. The database of language signals are record from actual lecturer and therefore sound more regular. However, the actual formant incidence values of the database of language signals are unknown, therefore; only qualitative analysis of the outcome can be make through diagram inspection. Algorithms are skilled for both male and female voice.

Formant frequencies of the backdrop speaker in its place of those of the main (more dominant) lecturer.

Table 1 value of formants in age group 15-25 male.

WORD	AB	AAP	ISE
FORMANT			
F1	522.8123455	507.449900	501.528774
F2	1416.166765	1439.56181	1574.94564
F3	2319.344543	2133.82727	2769.82366

Table.2 Values of formants in age group 15-25 female

WORD	AB	AAP	ISE
FORMANT			
F1	465.338826	421.760852	588.25110
F2	1486.51457	1794.760255	18968.7946
F3	2768.18918	2526.794171	3850.1362

It has been observed that in real-life, there is often more than just one orator there in an environs. The algorithm was also knowledgeable for the location in which surroundings speaker is there by estimate formant frequencies for the major lecturer in the presence the backdrop speaker. dissimilar cases are considered here. difficult is done with the female backdrop orator, with male setting speaker, with

multiple backdrop speaker. Here the background speaker serves while the 'sound source'. The volume of the backdrop speaker often vary in real-life . In some cases for a demanding time stage the background speaker may add vital energy to the formant incidence region of the primary professor. This will reason the algorithm to begin track .

VI. CONCLUSION

These algorithms are: Robust formant track algorithm formant tracking by RLS algorithm. Quantities learn of formant tracking algorithms have uncovered that it give precise formant frequency estimation for both male and female speaker for a broad variety in real-time noise situation such as many backdrop speaker Robust formant tracking algorithm provide classically flat formant frequency estimate than RLS algorithm. The robust formant tracking algorithm get well rapidly after mistaken approximation to go back to track the actual formant frequencies in the language signal, which is not the crate with RLS algorithm. since of this cause RLS algorithm show noisy tracking. in order concerning the gender is not available with RLS algorithm. But the computation complexity of RLS algorithm is less as evaluate to robust formant tracking algorithm. There have been some difficulty recognized with the robust formant tracker. The algorithm infrequently gives 'choppy' and oscillate formant frequency estimates. This is an surplus result because the real formant frequencies of words usually vary gradually with time and have smooth change. This difficulty is only encountered while the SNR is very short and occur due to the algorithm tracking the surplus energy added outside the formant frequency region from the backdrop sound source. However, in general production of the robust formant tracking algorithm is still a set recovered than that of formant tracking with RLS algorithm. In the training phase, each registered lecturer has to provide samples of their language so that the system can build or train a reference model for that lecturer. It consists of two main parts. The first part consists of processing each person's input voice sample to compress and summarize the characteristics of their vocal tracts. The second part involves pulling each person's data together into a single, easily manipulate matrix. The lecturer recognition system contains two main modules (i) feature extraction and (ii) feature matching. Feature removal is the process that extract a small amount of data from the voice signal that can later be used to represent each speaker.

VII.FUTURE WORK

The swing formant incidence difficulty might be solve in upcoming inform to the formant track algorithms by either flat the formant incidence estimate or by slot in extra rational limits to stop uneven jumps in the formant estimate.

VIII. REFERENCES

[1] Keller, E. 1994. *Fundamentals of Speech Synthesis and Speech Recognition*. Toronto: John Wiley & Sons

[2] Bazzi, I., Acero, A., and Deng, L. 2003. *An expectation maximization approach for formant tracking using a parameter-free non-linear predictor*, in Proceedings of ICASSP 2003, Hong Kong, pp. I.464–I.467

[3] Juang, B.-H., Chou, W., and Lee, C.-H. 1996. *Statistical and discriminative methods for speech recognition,* in *Automatic Speech and Speaker Recognition, Advanced Topics*, edited by C.-H. Lee, F. Soong, and K. Paliwal Kluwer Academic, Boston

[4] Weber, K. 2003. *HMM mixtures - HMM2 for robust speech recognition*, PhD thesis, Swiss Federal Institute of Technology Lausanne EPFL, Lausanne, Switzerland

[5] 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,

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

[7] 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

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

[9] 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.

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

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

[12] 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

[13] 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

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

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

[16] 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,

[17] 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

[18] 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,

[19] 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,

