iJRASET
International Journal For Research in Applied Science and Engineering Technology

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY

Volume: 3 **Issue:** VI **Month of publication:** June 2015

DOI:

www.ijraset.com

Call: 08813907089 | E-mail ID: ijraset@gmail.com

International Journal for Research in Applied Science & Engineering Technology (IJRASET)

An Overview on Speaker Identification Technologies

Shweta Bansal[1], Alok Kushwaha[2], S.S Agrawal[3]
[1,3]KIIT College of Engineering, Gurgaon, India,[2] Birla Institute of Technology Mesra, Muscat

Abstract - This paper aims at providing a brief overview into the area of speaker recognition. Speaker recognition can be classified into text dependent and the text independent methods. This paper gives an overview of major techniques developed in each stage of speaker recognition. This paper has a list of techniques along with their results, merits and demerits. After years of research and development the accuracy of speaker recognition remains one of the important research challenges (e.g., variations of the context, speakers, and environment).The design of Speaker Recognition system requires careful attentions to the following issues: feature extraction techniques, database and performance evaluation. The objective of this review paper is to summarize and compare some of the well known methods used in various stages of speaker recognition system and identify research topic and applications which are at the forefront of this exciting and challenging field.

I. INTRODUCTION

The speech signal contains many levels of information. Speech conveys the information about the language being spoken, the emotion, gender, and the identity of the speaker. The aim of automatic speaker recognition is to identity the speaker by extraction, characterization and recognition of the information contained in the speech signal. The area of speaker recognition is divided into two specific tasks i.e. verification and identification. In verification, the goal is to determine from a voice sample if a person is whom he or she claims. Generally it is assumed that person or user that falsely claiming to be a valid user are not known to the system, this task is referred as open set task. In speaker identification, the goal is to determine which one of a group of known voices best matches the input voice sample. It is assumed that unknown voice is coming from a fixed set of known speaker; this task is known as closed set identification.

There are various techniques and method s for speaker recognition. Researches are going on this area from last four decades and continue 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, Hidden Markov Models (HMMs) and Gaussian mixture model (GMM).The corpora for research and development in this area have evolved from small to large corpora. The applications of speaker recognition have been increasing since 1980's. [5]

The applications of speaker recognition technology use the speaker's voice for verification of their identity and thereafter enable the control access to services such as voice dialing and voice mail, tele-banking, telephone shopping, database access related services, information services, security control for confidential information areas, forensic applications, and remote access to computers. Speaker recognition is a commonly used biometric today. The background noise or the characteristic of communication channel can deteriorate the voice quality of the speaker therefore the speaker recognition system should be capable of accepting the wide range of variations in speaker's voice.

This paper reviews major highlights during the five decades in the research and development of speaker recognition system so as to provide a technological perspective. Although many technological progresses have been made, there still remain many research issues that need to be tackled.

II. BASIC STRUCTURE OF SPEAKER RECOGNITION SYSTEM

Speaker recognition systems generally consist of three major units as shown in figure1. The input to the first stage or the front end processing system is the speech signal. Here the speech is digitized and subsequently the feature extraction takes place. There are no exclusive features that convey the speaker's identity in the speech signal, however it is known from the source filter theory of speech production that the speech spectrum shape encodes in it the information about speaker's vocal tract shape via formants and glottal source via pitch harmonics. Therefore some form or the other of the spectral based features is used in most of the speaker

recognition systems. The process of speaker recognition consists of the training phase and the recognition phase. In the training phase, the features of a speaker's speech signal are stored as reference features. The feature vectors of speech are used to create a speaker's model. In the recognition phase, features similar to the ones that are used in the reference template are extracted from an input utterance of the speaker whose identity is required to be determined. The recognition decision depends upon the computed distance between the reference template and the template devised from the input utterance. In speaker identification, the distance between an input utterance and all of the available reference templates is computed. The template of the registered user, whose distance with the input utterance template is the smallest, is finally selected as the speaker of the input utterance. In case of speaker verification the distance is computed only between the input utterance and the reference template of the claimed speaker. If the distance is smaller than the predetermined threshold, the speaker is accepted other the speaker is rejected as an imposter [12].

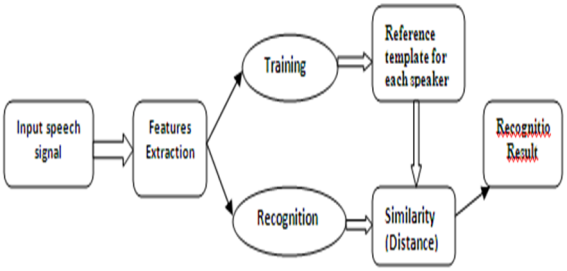


Figure1: Speaker Recognition system

III. FEATURE EXTRACTION TECHNIQUES

Speaker verification study was first conducted by Li *et al.* in 1966 using adaptive linear threshold elements [15]. This study used spectral representation of the input speech, obtained from a bank of 15 bandpass filters spanning the frequency range 300-4000 Hz. Two stages of adaptive linear threshold elements operate on the rectified and smoothed filter outputs. These elements are trained with fixed speech utterances. The training process results in a set of weights for the various frequency bands and time segments. The weights characterize the speaker. This study demonstrated that the spectral band energies as features contain speaker information. The study in [32] used pitch and formant information in addition to these band energies to improve the speaker verification performance. A study by Glenn *et al.* in 1967 suggested that acoustic parameters produced during the nasal phonation are highly effective for speaker recognition [33]. In this study, average power spectra of nasal phonation were used as the features for speaker recognition. In 1969, fast Fourier transform (FFT)-based cepstral coefficients were used in the speaker verification study [13]. In this work, a 34-dimensional vector was extracted from speech data. The first 16 components were from FFT spectrum, the next 16 were from log magnitude FFT spectrum and the last two components were related to pitch and duration. Such a 34-dimensional vector seems to provide a good representation of the speaker. A study made by G R Doddington in [14] reported an approach for speaker verification different from the approaches in [14] and [32]. He did not use a filter bank but converted the speech directly to pitch, intensity and formant frequency values, all sampled 100 times per second. These features were also demonstrated to provide good performance. Most of the above studies used spectral patterns of speech as features for speaker recognition. Atal in 1972 demonstrated the use of variations in pitch as a feature for speaker recognition [16]. In addition to variations in pitch, other acoustic parameters such as glottal source spectrum slope, word duration and voice onset time were proposed as features for speaker recognition by Wolf in 1971 [17]. The concept of linear prediction for speaker recognition was introduced by Atal in 1974 [34]. In this work, it was demonstrated that linear prediction cepstral coefficients (LPCCs) were better than the linear prediction coefficients (LPCs) and other features such as pitch and intensity. In general, the advantage of the cepstral coefficients is that they can be derived from a set of parameters which are invariant to any fixed frequency-response distortion introduced by the recording or transmission system [1].

Earlier studies neglected the features such as formant bandwidth, glottal source poles and higher formant frequencies, due to non-availability of measurement techniques. However, studies introduced after the linear prediction analysis, explored the speaker-specific potential of these features for speaker recognition [35]. A study carried out by Rosenberg and Sambur suggested that adjacent cepstral coefficients are highly correlated and hence all coefficients may not be necessary for speaker recognition [36]. In

1976, Sambur proposed to use orthogonal linear prediction coefficients as features in speaker identification [37]. In this work, he pointed out that for a speech feature to be effective, it should reflect the unique properties of the speaker's vocal apparatus and contain little or no information about the linguistic content of the speech. In 1977, long-term parameter averaging, which includes pitch, gain and reflection coefficients for speaker recognition, was studied [38]. In this study, it was shown that the reflection coefficients are highly informative and effective for speaker recognition. In 1981 Furui introduced the concept of dynamic features, to track the temporal variability in the feature vector in order to improve the speaker recognition performance [39][40]. A study by Reynolds in 1994 compared the different features like Mel frequency cepstral coefficients (MFCCs), linear frequency cepstral coefficients (LFCCs), LPCCs and perceptual linear prediction cepstral coefficients (PLPCCs) for speaker recognition [18]. He reported that among these features, MFCCs and LPCCs gave better performance than the other features. Though the MFCCs and LPCCs are used to extract the same vocal tract information, in practice these features differ in their performance due to the different principle involved in extracting it[12] , that is, the MFCC computation first applies discrete Fourier transform (DFT) on each frame and then weights the DFT spectrum by a Mel-scaled filter bank. The filter bank outputs are then converted to cepstral coefficients by applying the inverse discrete cosine transform (IDCT). In case of LPCCs, first, LPCs are obtained for each frame using Durbins-recursive method, and then these coefficients are converted to cepstral coefficients. Most of the studies discussed above considered vocal tract information as speaker characteristics for speaker recognition. In [41], it is reported that linear prediction (LP) residual also contains speaker-specific source information that can be used for speaker recognition. Also, it has been reported that though the energy of the LP residual alone gives less performance, combining it with LPCC improves performance as compared to that of the LPCC alone. On similar lines, several studies demonstrated that though the information from the LP residual alone gives less performance compared to the MFCC, combining it with MFCC improves the performance as compared to that of MFCC alone [20] [21] [24][25]. Recently, it has been reported that LP residual phase also contains speaker-specific source information [22]. In this study, it was demonstrated that the LP residual phase combined with MFCC improved the performance as compared to that of MFCC alone [22]. Plumpe *et al*. developed a technique for estimating and modeling the glottal flow derivative waveform from speech for speaker recognition. In this study, the glottal flow estimate was modeled as coarse and fine glottal features, which were captured using different techniques. Also, it was shown that the combined coarse and fine structured parameters gave better performance than the individual parameter alone [42]. Most of the studies discussed so far have not considered features like word duration, intonation, speaking rate, speaking style, etc., representing the behavioral traits, for speaker recognition. A study carried out in [43] demonstrated the significance of long-term pitch and energy information for speaker recognition. In another study, pitch tracks and local dynamics in pitch were also used in speaker verification [44]. A study in [45] reported that the combination of prosodic features like long-term pitch with spectral features provided significant improvement as compared to only the pitch features. A study carried out in [21] demonstrated the use of features like long-term pitch and duration information obtained using dynamic time warping (DTW), along with source and spectral features, for text-dependent speaker recognition. In [26], supra-segmental features like duration and intonation captured using neural networks were used for speaker recognition. In [46], amplitude modulation (AM)-frequency modulation (FM)-based parameters of speech were proposed for speaker recognition. In this study, it was demonstrated that using different instantaneous frequencies due to the presence of formants and harmonics in the speech signal, it is possible to discriminate speakers.

Among these, the mostly used ones are the spectral features, in particular, MFCCs and LPCCs. The main reasons for the same may be the less intra-speaker variability and also availability of rich spectral analysis tools. However, the speaker-specific information due to excitation source and behavioral trait represents different aspects of speaker information. Thus the feature extraction stage will benefit by using feature extraction techniques for excitation source and behavioral traits; however, the main limitation for the same is the non-availability of suitable tools for extracting the features, but this is where the future lies for the feature extraction stage.

IV. SPEAKER MODELING TECHNIQUES (YEAR WISE)

A. 1960-70

Template matching approach: In the direct template matching, training and testing feature vectors are directly compared using similarity measure. For the similarity measure, any of the techniques like spectral or Euclidean distance or Mahalanobis distance is used. Furui introduced the concept of dynamic time warping (DTW) for text-dependent speaker recognition [40]. However, it was

originally developed for speech recognition [47]. In this approach, the sequence of feature vectors of the training-speech signal is the text-dependent template model. The DTW finds the match between the template model and the input sequence of feature vectors from the testing-speech signal. The disadvantage of template matching is that it is time consuming, as the number of feature vectors increases. For this reason, it is common to reduce the number of training feature vectors by some modeling technique like clustering. The cluster centers are known as *codevectors*, and the set of codevectors is known as *codebook*. The most well-known codebook generation algorithm is the *K-means* algorithm [48, 49]. In 1985, Soong *et al. [50]* used the LBG algorithm for generating speaker-based vector quantization (VQ) codebooks for speaker recognition. It is demonstrated that larger codebook and larger test data give good recognition performance. Also, the study suggested that VQ codebook can be updated from time to time to alleviate the performance degradation due to different recording conditions and intra-speaker variations [50]. The disadvantage of the VQ classification is, it ignores the possibility that a specific training vector may also belong to another cluster.

The alternative of template-matching approach and VQ method for text-dependent speaker recognition is the HMM technique which was introduced in early 80's. HMM is the doubly stochastic process which as an underlying stochastic process that is not observable (hence the term hidden), but can be observed through another stochastic process that produces a sequence of observations. In HMM, time-dependent parameters are observation symbols. Observation symbols are created by VQ codebook labels. Continuous probability measures are created using Gaussian mixtures models (GMMs). The main assumption of HMM is that the current state depends on the previous state. In training phase, state transition probability distribution, observation symbol probability distribution and initial state probabilities are estimated for each speaker as a speaker model. The probability of observations for a given speaker model is calculated for speaker recognition. Kimbal *et al.* studied the use of HMM for text-independent speaker recognition under the constraint of limited data and mismatched channel conditions [58]. In this study, the MFCC feature was extracted for each speaker and then models were built using the broad phonetic category (BPC) and the HMM-based maximum likelihood linear regression (MLLR) adaptation technique.

B. 1990-2000

In 1995, Reynolds proposed Gaussian mixture modeling (GMM) classifier for speaker recognition task [69]. This is the most widely used probabilistic modeling technique in speaker recognition. The GMM needs sufficient data to model the speaker, and hence good performance. In the GMM modeling technique, the distribution of feature vectors is modeled by the parameters mean, covariance and weight. In another study, Reynolds compared GMM performance with regard to speaker identification with that of other classifiers like unimodal Gaussian, VQ, tied Gaussian mixture, and radial basis functions [71]. It was shown that GMM outperformed the other modeling techniques. Therefore, state-of-the-art speaker recognition systems use GMM as classifier due to the better performance, probabilistic framework and training methods scalable to large data sets [72]. The disadvantage of GMM is that it requires sufficient data to model the speaker well[70]. To overcome this problem, Reynolds *et al.* introduced GMM-universal background model (UBM) for the speaker recognition task [70]. In this system, speech data collected from a large number of speakers is pooled and the UBM is trained, which acts as a speaker-independent model. The speaker-dependent model is then created from the UBM by performing maximum *a posteriori* (MAP) adaptation technique using speaker-specific training speech. As a result, the GMM-UBM gives better results than the GMM. The advantage of the UBM-based modeling technique is that it provides good performance even though the speaker-dependent data is small. The disadvantage is that a gender-balanced large speaker set is required for UBM training.

As an alternative to the GMM, an auto-associative neural network (AANN) has been developed for pattern recognition task [61][73][74]. AANN is a feed-forward neural network which tries to map an input vector onto itself. The number of units in the input and output layers is equal to the size of the input vectors. The number of nodes in the middle layer is less than the number of units in the input or output layers. The activation function of the units in the input and output layer is linear, whereas the activation function of the units in the hidden layer can be either linear or nonlinear. The advantage of AANN over GMM is that, it does not impose any distribution. The application of AANN has been extensively studied for speaker recognition in [19][21][22][76]. A learning method based on the statistical learning theory, a special theory on machine learning, is the support vector machine (SVM). The SVM has many desirable properties, including the ability to classify sparse data without over-training. It is basically a solution to a two-class problem, but it can be extended to solve a multi-class problem by making it a one-versus-others two-class problem. SVM works by increasing the dimensionality of the input data space. The dimensionality is increased until it finds a maximum-margin linear hyperplane that can be used to separate the two classes. This is accomplished by using kernels and dot products. Moreover, SVM is discriminative in nature, whereas other classifiers are generative in nature. Vincent Wan and Steve Renals

studied SVM for speaker recognition [76, 77]. In these studies, different kernels, like the polynomial, the Fisher, a likelihood ratio and the pair HMM, were studied. It was reported that using these kernels it is indeed possible to achieve state-of-the-art speaker recognition performance. Further, the same authors have used score space kernels for speaker verification study in [77]. The score space kernels generalize Fisher kernels and are based on underlying generative models such as GMM. In this study, it was demonstrated that SVM reduced the error rate compared to GMM likelihood ratio system. In 2001, H Jiang and L Deng studied the Bayesian approach for speaker recognition [81]. It was demonstrated that Bayesian approach moderately improved the performance compared to well-trained baseline system using the conventional likelihood ratio test. In order to improve speaker recognition performance at the decision level, a combination of multiple classifiers has been proposed [82]. In this study, voting method was used for speaker identification based on the results of various resolution filter banks. A study conducted in [21] reported that by combining the evidences from source, supra-segmental and spectral features, it is indeed possible to improve the performance of the speaker recognition system. On similar lines, studies in [20, 22] have also demonstrated the combination of evidences from system and source features to improve performance. In [71] , it has been reported that the performance of the speaker recognition system can be improved by combining the evidences from SVM and GMM classifiers.

Recently a family of new normalization techniques has been proposed, in which the scores are normalized by subtracting the mean and then dividing by standard deviation, both terms having been estimated from the (pseudo) imposter score distribution. Different possibilities are available for computing the imposter score distribution: Znorm, Hnorm, Tnorm, Htnorm, Cnorm and Dnorm[3]. The state-of-the-art text-independent speaker verification techniques associate one or several parameterization level normalizations (CMS, feature variance normalization, feature warping, etc.) with word model normalization and one or several score normalizations. High-level features such as word idiolect, pronunciation, phone usage, prosody, etc. have been successfully used in text-independent speaker identification\verification. Typically, high-level-feature recognition systems produce a sequence of symbols from the acoustic signal and then perform recognition using the frequency and co-occurrence of symbols. In Doddington's work [13], word unigrams and bigrams from manually transcribed conversations were used to characterize a particular speaker in a traditional target/background likelihood ratio framework.

V. SUMMARY OF TECHNOLOGY PROGRESS

In the last 50 years, especially in the last three decades, research in speech 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 in the table 1[80].

S.No.	Past	Present
1	Template matching	Corpus-based statistical modeling, e.g. HMM and n grams
2	Filter bank/spectral resonance	Cepstral features, Kernel based function, group delay functions
3	Heuristic time normalization	DTW/DP matching
4	Distance -based methods	Likelihood based methods
5	Maximum likelihood approach	Discriminative approach e.g. MCE/GPD and MMI
6	Isolated word recognition	Continuous speech recognition,
7	Small vocabulary	Large vocabulary
8	Context Independent units	Context dependent units
9	Clean speech recognition	Noisy/telephone speech recognition
10	Single speaker recognition	Speaker-independent/adaptive recognition
11	Monologue recognition	Dialogue/Conversation recognition
12	Read speech recognition	Spontaneous speech recognition
13	Hardware recognizer	Software recognizer
14	Single modality(audio signal only)	Multimodal(audio/visual)speech recognition

Table 1: Progress in technologies in last 50 years

VI. CONCLUSION

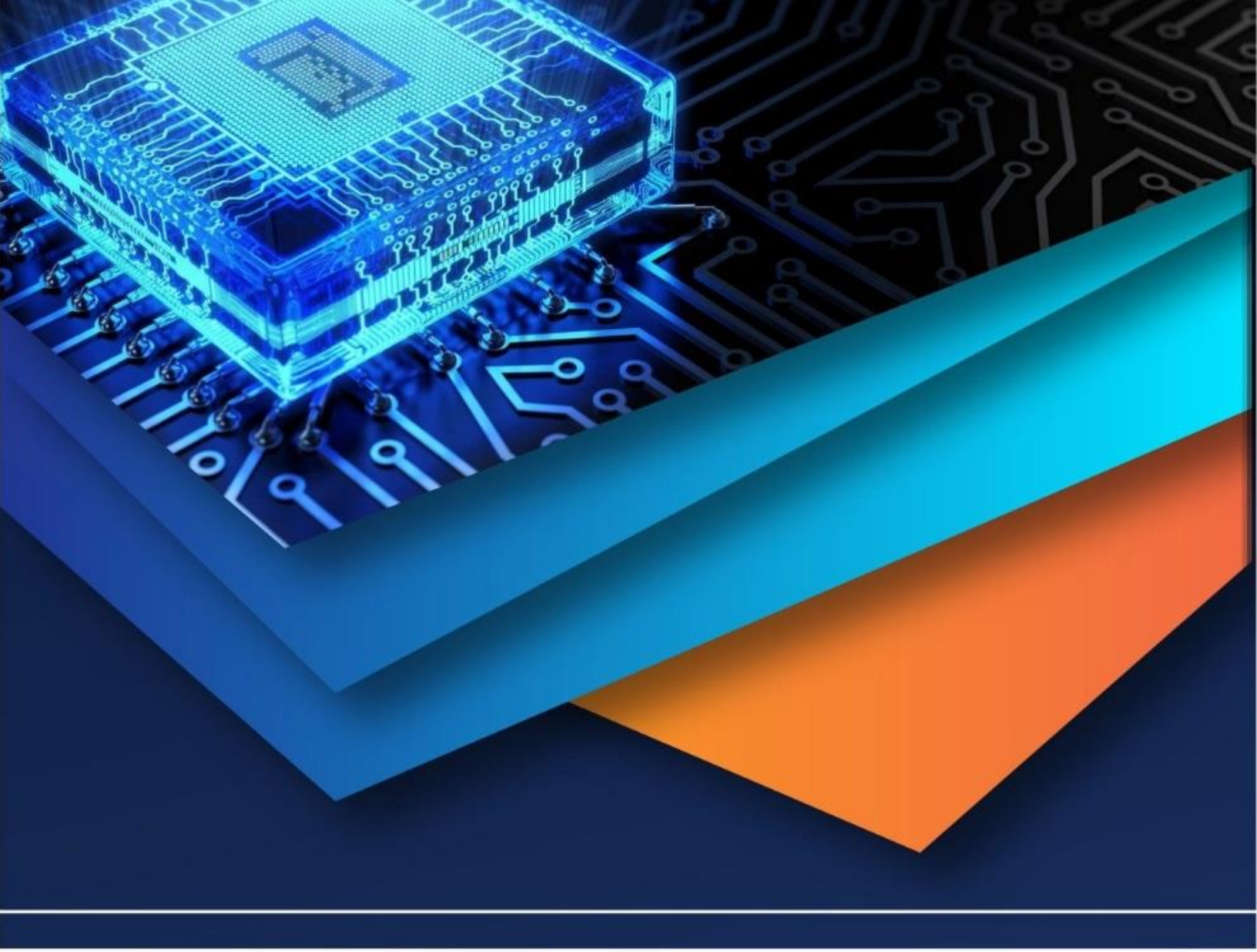
There has been a considerable amount of development in the field of speech and speaker recognition. The techniques for speaker recognition are yet to be successfully used in practical systems as the recognition rate is drastically reduced due to many reasons such as the distortion in the channel and the recording conditions and speaker-generated variability. Therefore it is important to explore stable features that remain insensitive to variation of speaker's voice over time and are robust against variation in voice quality due to colds or disguises. The problem of distortion in the channels and background noise also requires being resolved with better techniques. This paper attempts to provide a comprehensive survey of research on speaker recognition and to provide some year wise progress to this date. Although significant progress has been made in the last two decades, there is still work to be done.

REFERENCES

[1] B.S. Atal, "Automatic recognition of speakers from their voices, " Proc. IEEE , vol. 64(4), pp. 460-75, Apr. 1976.
[2]R.J. Mammone, X. Zhang, and R.P. Ramachandran, "Robust speaker recognition a feature-based approach, " IEEE Signal Process. Mag. , vol. 13(5), pp. 58-71, Sep. 1996.
[3]Piyush Lotia, "A Review of various score normalization techniques for speaker identification system, " Proce IJAET , May 2012.
[4]H. Gish, and M. Schmidt, "Text-independent speaker identification, " IEEE Signal Process. Mag. , vol. 18, pp. 18-32, Oct. 2002.
[5]J.P. Campbell, Jr., "Speaker recognition: A tutorial, " Proc. IEEE , vol. 85(9), pp. 1437-62, Sep. 1997.
[6]S.R. Mahadeva Prasanna, "Event based analysis of speech, " Ph.D. dissertation, Indian Institute of Technology Madras, Dept. of Computer Science, Chennai, India, Mar. 2004.
[7]P. Krishnamoorthy, "Combined temporal and spectral processing methods for speech enhancement, " Ph.D. dissertation, Indian Institute of Technology Guwahati, Dept. of Electronics and Communication Engg., Guwahati, India, Oct. 2008.
[8]P. Krishnamoorthy, and S.R.M. Prasanna, "Reverberant Speech Enhancement by Temporal and Spectral Processing, "IEEE Trans. Audio, Speech, Language Process. , vol. 17(2), p. 253-66, Feb. 2009.
[9]P.H. Arjun, "Speaker recognition in indian languages: A feature based approach, " Ph.D. dissertation, Indian Institute of Technology Kharagpur, Dept. of Electrical Engg., Kharagpur, India, Jul. 2005.
[10]G. Senthil Raja, "Feature analysis and compensation for speaker recognition under stressed condition, " Ph.D. dissertation, Indian Institute of Technology Guwahati, Dept. of Electronics and Communication Engg., Guwahati, India, Jul. 2007.
[11]V. Prakash, and J.H.L. Hansen, "In-Set/Out-of-Set speaker recognition under sparse enrollment, " IEEE Trans. Audio Speech Language Process. , vol. 15(7), pp. 2044-51, Sep. 2007.
[12]L. Rabiner, and B.H. Juang, Fundamentals of Speech Recognition. Singapore: Pearson Education, 1993.
[13] G. R. Doddington, "Speaker recognition based on
idiolectal differences between speakers," Proc.Eurospeech, pp. 2521-2524, 2001.
[14]G. Doddington, "Speaker recognition -identifying people by their voices, " Proc. IEEE , vol. 73, pp. 1651-64, 1985.
[15] K.P. Li, J.E. Dammann, and W.D. Chapman, "Experimental studies in speaker verification using an adaptive system, "J. Acoust. Soc. Amer. , vol. 40(5), pp. 966-78, Nov. 1966.
[16]B.S. Atal, "Automatic speaker recognition based on pitch contours, " J. Acoust. Soc. Amer. , vol. 52, no. 6(part 2), pp. 1687-97, 1972.
[17]J.J. Wolf, "Efficient acoustic parameters for speaker recognition, " J. Acoust. Soc. Amer. , vol. 51, no. 6(part 2), pp. 2044-56, 1971.
[18]D.A. Reynolds, "Experimental evaluation of features for robust speaker identification, " IEEE Trans. Speech Audio Process. , vol. 2(4), pp. 639-43, Oct. 1994.
[19]P. Satyanarayana, "Short segment analysis of speech for enhancement, " Ph.D. dissertation, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India , Feb. 1999.
[20]S.R.M. Prasanna, C.S. Gupta, and B. Yegnanarayana, "Extraction of speaker-specific excitation information from linear prediction residual of speech, " Speech Communication , vol. 48, pp. 1243-61, 2006.
[21]B. Yegnanarayana, S.R.M. Prasanna, J.M. Zachariah, and C.S. Gupta, , "Combining evidence from source, suprasegmental and spectral features for a fixed-text speaker verification system, " IEEE Trans. Speech Audio Process. , vol. 13(4), pp. 575-82, July 2005.
[22] K.S.R. Murthy, and B. Yegnanarayana, "Combining evidence from residual phase and MFCC features for speaker recognition, " IEEE Signal Process. Lett. , vol. 13(1), pp. 52-6, Jan. 2006.
[23] B. Yegnanarayana, K. Sharat Reddy, and S.P. Kishore, "Source and system features for speaker recognition using AANN models, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Utah, USA, Apr. 2001.
[24] K. Sharat Reddy, "Source and system features for speaker recognition, " Master's thesis, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India , 2001.
[25]C.S. Gupta, "Significance of source features for speaker recognition, " Master's thesis, Indian Institute of Technology Madras, Dept. of Computer Science and Engg., Chennai, India , 2003.
[26]L. Mary, K.S. Rao, S.V. Gangashetty, and B. Yegnanarayana, "Neural network models for capturing duration and intonation knowledge for language and speaker identification, " in Proc. Int. Conf. Cognitive Neural Systems, Boston, Massachusetts , May 2004.
[27]Farahani, P.G. Georgiou, and S.S. Narayanan, "Speaker identification using supra-segmental pitch pattern dynamics, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Montreal, Canada, May 2004, pp. 89-92.
[28] F. Weber, L. Manganaro, B. Peskin, and E. Shriberg, "Using prosodic and lexical information for speaker identification, " in proc. Int. Conf. Acoust., Speech, Signal Process. , vol. 1, London, UK, April. 2002, pp. 141-4.

[29]P. Denes, and M.V. Mathews, "Spoken digit recognition using time-frequency pattern matching, " J. Acoust. Soc. Amer. , vol. 32(11), pp. 1450-5, Nov. 1960.
[30]S. Pruzansky, "Pattern-matching procedure for automatic talker recognition, " J. Acoust. Soc. Amer. , vol. 35(3), pp. 354-8, Mar. 1963.
[31]S. Pruzansky, and M.V. Mathews, "Talker-recognition procedure based on analysis of variance, " J. Acoust. Soc. Amer., vol. 36(11), pp. 2041-7, Nov. 1964.
[32]S.K. Das, W.S. Mohn, and S.L. Saleeby, "Speaker verification experiments, " J. Acoust. Soc. Amer. , vol. 49, p. 138(A), 1971.
[33]J.W. Glenn, and N. Kleiner, "Speaker identification based on nasal phonation, " J. Acoust. Soc. Amer. , vol. 43(2), pp. 368-72, June 1967.
[34]B.S. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification, " J. Acoust. Soc. Amer. , vol. 55, pp. 1304-12, 1974.
[35] M.R. Sambur, "Selection of acoustic features for speaker identification, " IEEE Trans. Acoust., Speech, Signal Process. , vol. ASSP-23(2), pp. 176-82, Apr. 1975.
[36]A.E. Rosenberg, and M.R. Sambur, "New techniques for automatic speaker verification, " IEEE Trans. Acoust., Speech, Signal Process. , vol. ASSP-23(2), pp. 169-76, Apr. 1975.
[37] M.R. Sambur, "Speaker recognition using orthogonal linear prediction, " IEEE Trans. Acoust., Speech, Signal Process.. , vol. ASSP-24(4), pp. 283-9, Aug. 1976.
[38]J.D. Markel, B.T. Oshika, and A.H. Gray, Jr., "Long-term feature averaging for speaker recognition, " IEEE Trans. Acoust., Speech, Signal Process. , vol. ASSP-25(4), pp. 330-7, Aug. 1977.
[39] S. Furui, "Speaker-independent isolated word recognition using dynamic features of speech spectrum, " IEEE Trans. Acoust., Speech, Signal Process. , vol. ASSP-34, pp. 52-9, Feb. 1986.
[40]Sasaoki Furui, "Cepstral analysis technique for automatic speaker verification, " IEEE Trans. Acoust., Speech, Signal Process. , vol. 29(2), pp. 254-72, Apr. 1981.
[41] P. Thevenaz, and H. Hugli, "Usefulness of the LPC-residue in text-independent speaker verification, " Speech Communication , vol. 17, pp. 145-57, 1995.
[42] M.D. Plumpe, T.F. Quatieri, and D.A. Reynolds, "Modeling of the glottal flow derivative waveform with application to speaker identification, " IEEE Trans. Speech Audio Process. , vol. 7(5), pp. 569-85, 1999.
[43]M.J. Carey, E.S. Parris, H. Lloyd-Thomas, and S. Bennett, "Robust prosodic features for speaker identification, " inproc. Int. Spoken Language Process. , Philadelphia, PA, USA, Oct. 1996.
[44]M.K. Sonmez, E. Shriberg, L. Heck, and M. Weintraub, "Modeling dynamic prosodic variation for speaker verification, " in proc. Int. Spoken Language Process. , Sydney, NSW, Australia, Nov-Dec. 1998.
[45] B. Peskin, J. Navratil, J. Abramson, D. Jones, D. Klusacek, D.A. Reynolds, and B. Xiang, "Using prosodic and conversational features for high-performance speaker recognition, " in Int. Conf. Acoust., Speech, Signal Process. , vol. IV, Hong Kong, Apr. 2003, pp. 784-7.
[46]M. Grimaldi, and F. Cummins, "Speaker identification using instantaneous frequencies, " IEEE Trans. Audio, Speech, Language Process. , vol. 16(6), pp. 1097-111, Aug. 2008.
[47]H. Sakoe, and S. Chiba, "Dynamic programming algorithm optimization for spoken word recognition, " IEEE Trans. Acoust., Speech, Signal Process. , vol. 26, pp. 43-9, Feb. 1978.
[48] Y. Linde, A. Buzo, and R.M. Gray, "An algorithm for vector quantizer design, " IEEE Trans. Communications , vol. COM-28(1), pp. 84-96, Jan. 1980.
[49] R. Gray, "Vector quantization, " IEEE Acoust., Speech, Signal Process. Mag. , vol. 1, pp. 4-29, Apr. 1984.
[50] F.K. Soong, A.E. Rosenberg, L.R. Rabiner, and B.H. Juang, "A Vector quantization approach to speaker recognition, " in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. , vol. 10, Detroit, Michingon, Apr. 1985, pp. 387-90.
[51]J.C. Bezdek, and J.D. Harris, "Fuzzy portions and relations;an axiomatic basis for clustering, " Fuzzy Sets and Systems , vol. 1, pp. 111-27, 1978.
[52]H.J. Zimmermann, Fuzzy set theory and its applications , 1st ed. Kluwer academic, 1996.
[53]L. Lin, and S. Wang, "A Kernel method for speaker recognition with little data, " in Int. Conf. signal Process. , Budapest, Hungery, May, 2006.
[54]V. Chatzis, A.G. Bors, and I. Pitas, "Multimodal decision-level fusion for person authentication, " IEEE Trans. Man Cybernetics Part A: Systems and Humans , vol. 29, pp. 674-81, Nov. 1999.
[55]A.E. Rosenberg, and S. Parthasarathy, "Speaker background models for connected digit password speaker verification, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Atlanta Georgia, May 1996, pp. 81-4.
[56]J.M. Naik, L.P. Nestch, and G.R. Doddington, "Speaker verification using long distance telephone lines, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Glasgow, UK, May 1989, pp. 524-7.
[57]T. Matsui, and S. Furui, "Comparison of text-independent speaker recognition methods using VQ-distortion and Discrete/continuous HMMs, " IEEE Trans. Speech Audio Process. , vol. 2(3), pp. 456-9, July 1994.
[58]O. Kimball, M. Schmidt, H. Gish, and J. Waterman, "Speaker verification with limited enrollment data, " in proc. European Conf. Speech Commun. and Tech. (EUROSPEECH'97) , Rhodes, Greecee, Sep. 1997, pp. 967-70.
[59]R.P. Lipmann, "An introduction to computing with neural nets, " IEEE Trans. Acoust., Speech, Signal Process. , vol. 4, pp. 4-22, Apr. 1989.
[60] G. Bannani, and P Gallinari, "Neural networks for discrimination and modelization of speakers, " Speech Communication. , vol. 17, pp. 159-75, 1995.
[61]B. Yegnanarayana, Artificial neural networks. New Delhi: Prentice-Hall, 1999.
[62]J. Oglesby, and J.S. Mason, "Optimization of neural models for speaker identification, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Albuquerque, NM, May 1990, pp. 261-4.
[63]T. Kohonen, "The self-organizing map," Proce. IEEE , vol. 78(9), pp. 1464-80, Sep. 1990.
[64]M. Inal, and Y.S. Fatihoglu, "Self organizing map and associative memory model hybrid classifier for speaker recognition, " in proc. Neu., Net., App., Elec., Engg. (NEUREL'02) , Belgrade, Yugoslavia, Sep. 2002, pp. 71-4.
[65]A.T. Mafra, and M.G. Simoes, "Text independent automatic speaker recognition using self-organizing maps, " in proc. Ind. App. Society conf. , vol. 3, Victoria, British Columbia, Oct. 2004, pp. 1503-10.
[66] G. Bannani, F. Fogelman, and P. Gallinari, "A connectionist approach for speaker identification, " in proc. Int. Conf. Acoust., Speech, Signal Process. , Albuquerque, NM, May 1990, pp. 265-8.

[67]J. He, L. Liu, and G. Palm, "A discriminative training algorithm for VQ-based speaker identification, " IEEE Trans. Speech Audio Process. , vol. 7, pp. 353-6, May 1999.
[68]D.A. Reynolds, "Speaker identification and verification using Gaussian mixture speaker models," Speech Communication, vol. 17, pp. 91-108, 1995.
[69 D.A. Reynolds, and R.C. Rose, "Robust text-independent speaker identification using Gaussian mixture speaker models, " IEEE Trans. Speech Audio Process. , vol. 3, pp. 72-83, Jan. 1995.
[70] W.M. Campbell, J.P. Campbell, D.A. Reynolds, E. Singer, and P.A. Torres-Carrasquillo, "Support vector machines for speaker and language recognition, " Computer Speech and Language , vol. 20, pp. 210-29, 2006.
[71] D.A. Reynolds, T.F. Quateri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models, " Digital Signal Processing , vol. 10, pp. 19-41, 2000.
[72]B. Yegnanarayana, and S.P. Kishore, "AANN: An alternative to GMM for pattern recognition, " Neural Networks, vol. 15, pp. 459-69, 2002.
[73] M. Shajith Iqbal, Hemanth Misra, and B. Yegnanarayana, "Analysis of auto associative neural networks, " in Int. Joint Conf. Neural Networks , Washington, USA, 1999.
[74] B. Yegnanarayana, K.S. Redddy, and S.P. Kishore, "Source and system features for speaker recognition using AANN models, " in Int. Conf. Acoust., Speech, Signal Process. , Salt Lake City, Utah, USA, Apr. 2001, pp. 409-12.
[75]N. Dhananjaya, and B. Yegnanarayana, "Correlation-based similarity between signals for speaker verification with limited amount of speech data, " in proc. International Workshop, MRCS 2006 , Istanbul, Turkey, Sep. 2006.
[76]V. Wan, and S. Renals, "Speaker verification using sequence discriminant support vector machines, " IEEE Trans. Speech Audio Process. , vol. 13, pp. 203-10, 2005.
[77]V. Wan, and S. Renals, "Evaluation of kernel methods for speaker verification and identification, " in Proc. IEEE Int. Conf. Acoust., Speech, Signal Process. , vol. 1, pp. 1-669 - 1-672, 2002.
[78]W.M. Campbell, D.E. Sturim, and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification, " IEEE Signal Process. Lett. , vol. 13(5), pp. 308-11, May 2006.
[79]C.H. You, K.A. Lee, and H. Li, "An SVM kernel with GMM-supervector based on the Bhattacharyya distance for speaker recognition, " IEEE Signal Process. Lett. , vol. 16(1), pp. 49-52, Jan. 2009.
[80]G.R. Doddington, M.A. Przybocki, A.F. Martin, and D.A. Reynolds, "The NIST speaker recognition evaluation overview, methodology, systems, results, perspective, " Speech Communication , vol. 31, pp. 225-54, 2000.
[81]H. Jiang, and L. Deng, "A Bayesian approach to the verification problem: Applications to speaker verification, " IEEE Trans. Speech, Audio Process. , vol. 9(8), pp. 874-975, 2001.
[82]B-J Lee, S-W Yoon, H-G Kang, and D.H. Youn, "On the use of voting methods for speaker identification based on various resolution filterbanks, " in proc. Int. Conf. Acoust., Speech, Signal Process. , vol. I, Toulouse, France, May 2006, pp. 917-20.



10.22214/IJRASET

45.98

IMPACT FACTOR: 7.129

IMPACT FACTOR: 7.429

INTERNATIONAL JOURNAL FOR RESEARCH

IN APPLIED SCIENCE & ENGINEERING TECHNOLOGY