International Journal of Recent Technology and Engineering (IJRTE)
ISSN: 2277-3878 (Online), Volume-7 Issue-4s2, December 2019

A Comparison of Features for Multilingual Speaker Identification - A Review and Some Experimental Results

Pritam Limbaji Sale, Spoorti J Jainar, B.G. Nagaraja

Abstract: ***Countries like India, Canada, Malaysia, etc. are multilingual in nature. People in multilingual countries have habituated to use several languages. Due to the increased number of multilingual speaker identification system applications, the interest in the area has grown notably in recent years. The accuracy of speaker recognition system is severely degraded if training and testing speech languages are different. In speaker recognition area, researchers have made many attempts to tackle language mismatch issues. Choosing a suitable feature extraction method for obtaining appropriate information using speech signal is an essential task. This paper reports a concise experimental review of ten feature extraction techniques for the multilingual scenario. The monolingual, crosslingual and multilingual speaker identification studies are carried out using randomly selected 50 speakers from the IITG multi-variability speaker recognition (IITG-MV) database. Comparative results indicate that subband centroid frequency coefficients (SCFC), linear frequency cepstral coefficients (LFCC) and multitaper Mel frequency cepstral coefficients (MFCC) features are considerably more useful in all the speaker identification. Further, concluding any relation to speaker identification performance in the language mismatch environment is identification as the distribution of speakers in different languages is non-uniform.***

Keywords: Speaker identification, monolingual, cross lingual, multilingual, LPCC, MFCC, IMFCC, LFCC, RFCC, multitaper MFCC, SCFC, SSFC, GMM-UBM.

I. INTRODUCTION

Speaker recognition is one among the important problems in speech signal processing. Speaker recognition system, carry out either speaker verification or speaker identification tasks [1], [2]. The speaker identification task results in the best match for an unknown voice from the database, whereas speaker verification task is either to accept/reject a claimed identity. Based on the languages used for the study, speaker recognition can be done in monolingual,crosslingual and multilingual modes [3]. In the monolingual mode, same language is considered for both training and testing. In crosslingual speaker recognition, training and testing are carried out with different languages. In multilingual mode, speaker-dependent models are trained using single language and tested for several languages [3], [4]. Monolingual and crosslingual speaker recognitions are a subset of multilingual speaker recognition.

In [5], speaker discriminative power of Mel frequency cepstral coefficients (MFCC), anti-MFCC, and linear frequency cepstral coefficients (LFCC) was examined using telephone speech data. It was demonstrated that LFCCs perform considerably better than MFCCs for speaker recognition systems using the nasal and non-nasal consonant broad phonetic regions. MFCC performs better in the lower frequency region, but in the higher frequency regions, it does not capture speaker characteristics effectively and hence its performance degrades [6].

A new speaker recognition system was introduced based on MFCC and back-propagation neural networks [6]. It was observed from the results that proposed system was feasible and fairly successful. A robust speaker recognition system based on multi-stream features (spectral centroid frequency and power normalized cepstral coefficients) was proposed in [7]. Three types of fusion technique were introduced and com-pared. Results proved the effectiveness of proposed multi-stream features for speaker recognition system. In [8], a new type of feature was proposed based on gammatone frequency cepstral coefficients (GFCC) by introducing technologies like multitaper estimation, smooth amplitude envelope, mean subtraction, variance normalization and autoregressive moving average filter. The experimental results using TIMIT database showed that under different noise and different signal-to-noise ratio, the proposed GFCC performs better than the MFCC.

For multilingual countries, the impact of different languages on speaker recognition system needs to be studied. Non-availability of the standard multilingual database for Indian languages is the major difficulty for carrying out speaker recognition research in multilingual/language mismatched scenario. In this direction, a study on the effect of language mismatch on speaker identification and speaker verification task was carried out in [9]. The speech data (100 speakers) was recorded in three languages, viz., the local language of Arunachal Pradesh, Hindi, and Indian English. It was studied that the impact of language variability in intra-group and inter-group were nearly same. A study in [10], focused on the speaker recognition performance under language mismatch environment.

Further, the consequence of language mismatch was also described for a single language and multiple languages together.

Manuscript published on 30 December 2019.
***Correspondence Author(s)
Pritam Limbaji Sale, Research Scholar, VTU, Belagavi.
Spoorti J. Jainar, Research Scholar, VTU, Belagavi.
B.G. Nagaraja, Professor & Head, Dept. of E&CE, JIT, Davangere.

The experiments were demonstrated using 13 Indian languages. It was observed that the impact of language mismatch was less significant in speaker identification task compared to speaker verification. Literature reveals that the most of the investigations are restricted to a single language environment. It is essential to know which type of feature is more beneficial for a multilingual speaker recognition. In this work, we provide a brief overview of the feature extraction techniques for speaker identification systems. Following this overview, we will experimentally compare 10 front-end speech features for multilingual speaker identification using Gaussian mixture model with universal background model (GMM-UBM).

In the rest of this paper, Section 2 describes the different front-end feature extraction techniques for speaker recognition. A database used for the study and the experimental setups are provided in Section 3. Section 4 discusses the monolingual, crosslingual and multilingual speaker identification results. The paper concludes in Section 5.

II. FEATURE EXTRACTION TECHNIQUES

2.1 Linear prediction cepstral coefficients (LPCC)

The cepstral features obtained from either filter-bank approach or a linear prediction analysis are relatively treated as state-of-the-art front-end features for speaker identification. In linear prediction (LP) speech analysis, every sample is approximated as a linear weighted sum of the past speech samples. In LPCC, the power spectrum is computed from the smoothed auto-regressive power spectral estimate instead of the periodogram estimate technique. The order of the linear prediction analysis denotes the number of peaks in an all-pole model. The predictor coefficientsa_k are transformed into robust cepstral coefficients. The relationship between predictor coefficients, a_k and cepstral coefficients, C_n is characterized by the following recursive relation [11] given as:

$$C_o = ln\, G \qquad (1)$$

$$C_n = a_n + \sum_{k=1}^{n-1} \left(\frac{k}{n}\right) c_k a_{n-k}\ ;\ 1 \le n \le p \qquad (2)$$

Where$G$ is the gain factor in the LP model, n is an index and p is the LP order.

2.2 MFCC

The MFCC feature extraction method consists of windowing (Hamming) the signal to minimize the spectral distortion, applying the discrete Fourier transform (DFT) to obtain the magnitude frequency response. To compute the log energy, frame magnitude spectra is multiplied with the set of Mel-scale triangular filters. Finally, inverse discrete cosine transform of filter bank coefficients is calculated to get the cepstral coefficients. The mapping of perceived frequency (f_{mel} in mels) with respect to the physical frequency (f in Hz) scale is given by [12]:

$$f_{mel} = 2595\ log_{10}\left(1 + \frac{f}{700}\right) \qquad (3)$$

2.3 Sine-weighted cepstrum estimator (SWCE) multitaper MFCC

From a statistical perspective, the MFCC feature extraction using windowed DFT is not favorable due to large variance of the spectrum approximation [13]. Efforts have been made to improve the MFCC robustness. A study in [13]-[15] described the multitaper MFCC feature extraction technique for speaker recognition. In order to lower the variance of the MFCC estimator, multiple window (multitaper) spectrum estimates was used. In multitapering technique, analysis frame is passed through different window functions and the final spectrum estimate was computed using a weighted mean of the individual sub-spectra. Every individual spectral contribute to a final spectral envelope for each component.

Thomson multitaper [16], multipeak multitaper [17] and SWCE multitaper [18] are the popular multitapering techniques. In [13], it was described that the choice of the number of tapers K was more important than the choice of multitaper type. Further, the best results were obtained for $3 \le K \le 8$. Our previous results showed that the SWCE multitaper MFCC ($K = 6$) can be used to improve the speaker identification performance for language mismatched scenario [4]. In this work, SWCE multitapers are used with $K = 6$ windows. The weights employed in the SWCE multitaper method are described by the following closed-form expression [13]:

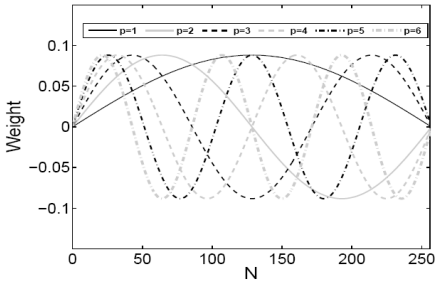

Figure 1: SWCE method for six sine tapers, p is the taper number and N = 256 window length.

$$\lambda(p) = \frac{cos\left(\frac{2\pi(p-1)}{K/2}\right)+1}{\sum_{p=1}^{K}\left(cos\left(\frac{2\pi(p-1)}{K/2}\right)+1\right)} : p = 1 \dots \dots \dots \dots K \qquad (4)$$

The Figure 1 shows the sine tapers of the SWCE multitaper method for $K = 6$.

2.4 τth - order MFCC

A novel type of windowing technique to extract MFCC features for speaker recognition task was presented in [19]. The new window technique uses differentiation in the frequency domain property of discrete time Fourier transforms. Let $\omega(n)$ be the Hamming window function, the proposed window function ($\omega(n)$) for τ^{th} - order is obtained as:

$$\hat{\omega}(n) = n^{\tau}\omega(n) \qquad (5)$$

Then windowed speech frame is described as:

$$\hat{x}(n) = \hat{\omega}(n)x(n) \qquad (6)$$

where$x(n)$ is a raw speech frame. The Hamming window can be analyzed as zero order window ($\tau = 0$) of proposed family. Figure 2 shows the frequency and time domain representation of $\hat{\omega}$(n) window function for 160 samples. In this work, τ^{th} - order MFCC are computed for $\tau = 2$.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

2.5 Inverted Mel frequency cepstral coefficients (IMFCC)

In [12], IMFCC features were proposed for speaker identification. IMFCC features efficiently capture high-frequency formants, which are neglected by the conventional MFCC. In IMFCC, the complete filter bank structure is inverted so that

the upper-frequency width averages by more precisely located filters and in the lower frequency range lesser number of widely located filters are utilized. Inverted Mel scale ($\hat{f}_{mel}$) is defined as [12]:

$$\hat{f}_{mel} = 2195.2860 - 2595\, log_{10}\left(1 + \frac{4031.25-f}{700}\right) \quad (7)$$

2.6 LFCC

Based on the theory of human speech production, several speakers specific information connected with the vocal tract structure, mainly, length of the vocal tract are distributed more in the higher frequency region. Hence, linear frequency scale may offer various benefits in speech/speaker recognition in comparison with the conventional Mel frequency scale. In this direction, a study in [20] compared the LFCC and MFCC features using the NIST speaker recognition evaluation (SRE) 2010 extended-core task database. Results using SRE10 showed that the LFCC constantly perform better than the MFCC due to its higher recognition for the female subjects. The feature extraction methodology of LFCC is almost similar to MFCC. The only difference is that the Mel-scale filters are replaced by linear space triangular filter bank [21].

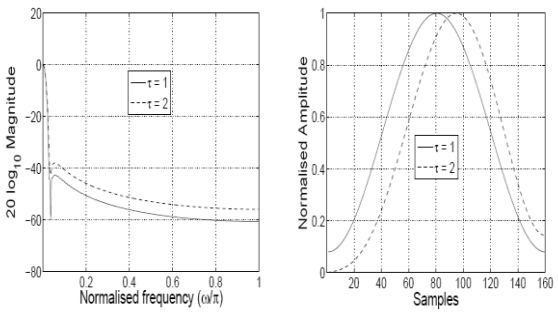

Figure 2: Frequency and time domain representation of $\hat{\omega}$(n) window function for 160 samples.

2.7 Rectangular filter-bank cepstral coefficient (RFCC)

RFCC feature was originally proposed for robust automatic speech recognition in noisy/Lombard speech conditions [22]. In [22], RFCC features were used as one of the front end features for the 2012 NIST SRE. It was studied that the overall recognition performance was very successful. The RFCC feature extraction technique is inspired by perceptual linear prediction cepstral features [22]. In RFCC extraction, integration is carried out using a rectangular window and the filters spaced in linear scale [21].

2.8 All-pole group delay function (APGDF)

The significance of phase-based features for speaker verification was investigated in [23]. In that work, group delay function was used to derive the speaker-specific features from the phase spectrum. It was mentioned that the group delay features using all-pole models can be utilized to compute phase information for speaker verification task. Speaker verification results using NIST 2010 SRE dataset showed that the group delay techniques are useful under the high vocal effort. Further, group delay features were comparable to usual magnitude spectrum-based MFCC features. The group delay function $\tau(x)$ for the speech signal $s(n)$ can be obtained as derived in [23], [24]:

$$\tau(w) = \left(\frac{S_R(w)X_R(w) + S_I(w)X_I(w)}{|S(w)|^2}\right) \quad (8)$$

where $S(w)$ is a Fourier transform of $S(n)$, $X(w)$is a Fourier transform of $x(n)$,$x(n)= ns(n)$, $|S(w)|$ is the magnitude spectrum, $S_R(w)$ is a real part of $S(w)$,$S_I(w)$ is an imaginary part of $S(w)$, $X_R(w)$ is a real part of $X(w)$and $X_I(w)$ is an imaginary part of $X(w)$.

2.9 Sub band centroid frequency coefficients (SCFC)

Attempts have been made to obtain the alternative to the cepstral features that should provide information for speech/speaker recognition. In this direction, Paliwal [25], investigated a few formant type features for speech recognition. In [26], SCF was used for text-independent speaker authentication as a complement to cepstral based features. It was concluded that spectral subband centroids (SSC) perform somewhat better than MFCC features and also combining SSC with MFCC features enhance the performance of the authentication system in both clean and noisy environments. Further, SCFC includes complementary knowledge related to subbands which are not present in cepstral features.

Let $S[n]$ represent a speech frame, for n = 0,1,…,N - 1, and $S[f]$ correspond to the speech frame spectrum. Further, $S[f]$ is split into K number of subbands and each subband is described as upper frequency edge (u_k) and lower frequency edge (l_k). $W_k\,[f]$Denotes the frequency-sampled response of the filter. SCF defined as the weighted average frequency for a given subband [27]. The k^{th}sub band spectral centroid frequency F_k described as [25]-[27]:

$$F_k = \frac{\sum_{f=l_k}^{u_k} f|S\,[f]W_k\,[f]|}{\sum_{f=l_k}^{u_k}|S[f]W_k[f]|} \quad (9)$$

2.10 Sub band spectral flux coefficients (SSFC)

In [28], spectral flux based features were used for speech and music discrimination task. Spectral flux determines the frame-to-frame variation of speech in the power spectrum [28]. It is computed between normalized power spectrum of consecutive frames using the Euclidean distance measure [21].

Recently, a new feature called SSFC features are used for synthetic speech detection. It was mentioned that the SSFC features contain information connected to spectral flux in different subbands.

Experimental results showed that the SSFC features were useful compared to other spectral features considered for the study. The subband spectral flux (SSF) of the i^{th}subband of the m^{th} speech frame is [21]:

$$SSF_m^i = \sum_{k=1}^{N/2+1}\|\bar{S}_m(k) - \bar{S}_{m-1}(k)\|^2 W_i(k) \quad (10)$$

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

where $W_i(k)$ denotes the spectral window function used to find the frequency response of i^{th}subband, $\bar{S}_m(k)$ is the magnitude of k^{th} frequency element of the normalized power spectrum of m^{th} frame and N is the number DFT bins.

Then SSFCs are obtained by applying logarithmic compression and DCT on SSFs.

III. DATABASE AND EXPERIMENTAL SETUP

All the experiments are carried out using the IITG multi-variability speaker recognition (IITG-MV) database. The speech data was collected using different speaking styles, sensors, environments and languages from 200 speakers [29].

The recording was carried out in the wide varieties of acoustic environments (office, corridors, hostel rooms, laboratories, etc.). The speech data are collected in multi-sensor (five different sensors), multilingual (any of the favorite Indian language and Indian English) and multi-style (reading and conversational styles) conditions. The languages include Indian English, Hindi, Tamil, Kannada, Oriya, Malayalam, Assami and so on [30]. The speech signal was sampled at 8000 Hz and 16 bits depth. For this work, 50 speakers (20 female and 30 male) in the IITG-MV database is used.The UBM was built using diagonal covariance matrices. To build the gender independent UBM, first 18 speech _les from enroll data of all the speakers (nearly two hours) of YOHO database [31] is used. The GMM model parameters (mean vector, mixture weights, and covariance matrix) were measured using the expectation maximization iterative algorithm. Maximum a posteriori technique is used to build the speaker-dependent models (GMM) by considering only the mean parameters of the UBM [29]. Speakers are modeled using GMMs with 16, 32, 64, and 128 mixtures.

The number of speakers (50) and speech duration (20 s) is kept constant throughout the experiment to perform a relative comparison of monolingual, crosslingual and multilingual speaker identification results using different feature extraction techniques.

For feature extraction, analysis window of 20 ms duration with 10 ms intervals was considered. The dimension of all the features is 13 (excluding 0^{th} coefficient).

IV. SPEAKER IDENTIFICATION RESULTS

In this section, the speaker identification results are discussed in different languages, viz., monolingual, crosslingual and multilingual.

The best results in each condition (P_i) are highlighted.

Note: A/B indicates training in language ‘A’ and testing with language `B'; `multi-language' includes Kannada, Assami, Bengali, Telugu, Tamil, Malayalam and Oriya.

Table 1-3 give the performance of LPCC, MFCC, multitaper MFCC, τ^{th}- order MFCC, IMFCC, LFCC, RFCC, APGDF, SCFC and SSFC for monolingual, crosslingual and multilingual speaker identification.

Some of the observations from the speaker identification experiments are as follows:

Table 1: The monolingual speaker identification performance using different features and GMM-UBM classifier.

Training/ Testing	Features	Gaussian mixtures				P_i
		16	32	64	128	
E/E	LPCC	76	74	84	86	86
	MFCC	64	74	78	84	84
	multitaper MFCC	72	78	82	82	82
	τ^{th} - order MFCC	70	76	82	80	82
	IMFCC	60	64	64	74	74
	LFCC	66	70	76	84	84
	RFCC	62	70	72	78	78
	APGDF	68	68	78	76	78
	SCFC	72	84	86	86	86
	SSFC	64	68	76	78	78
H/H	LPCC	70	74	80	82	82
	MFCC	74	74	78	80	80
	multitaper MFCC	78	78	84	82	84
	τ^{th} - order MFCC	70	74	80	82	82
	IMFCC	62	68	68	76	76
	LFCC	76	80	78	84	84
	RFCC	66	70	78	80	80
	APGDF	64	72	78	76	78
	SCFC	72	78	82	82	82
	SSFC	70	76	80	76	80

- The results indicate that the recognition performance of 86% is achieved for LPCC and SCFC features when training and testing are done in English language (E/E).
- The multitaper MFCC and LFCC features give the highest performance of 84% when training and testing are done in Hindi language (H/H).
- The results indicate that the recognition performance of 80% is achieved for multitaper MFCC, τ^{th}- order MFCC and SCFC features when training is done in English language and tested with Hindi language (E/H).
- The SCFC features give the highest performance of 78% speaker when training is done in Hindi and tested with the English language (H/E).
- The results show that the recognition performance of 82% is achieved for τ^{th} - order MFCC, LFCC and SCFC features when the training is done in English and testing with multi language (E/multi).
- The LPCC features give the highest performance of 82% when training is done in Hindi and testing with multi language (H/multi).
- In the majority of the cases, monolingual speaker identification performs better as compared to crosslingual and multilingual scenario. These results clearly indicate that the effect of languages on speaker identification.
- The performance of multilingual based speaker identification system is better than that of the crosslingual, but poor compared to the monolingual system.
- The recognition performance significantly degraded in crosslingual may be due to the deviation in fluency and stressing of few words when the same person speaks different languages.
- Further, due to dissimilar prosodic and phonetic models of the languages [32].
- The SCFC performs better in the majority of the speaker identification experiments.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

- The speaker recognition performance using LFCC and multitaper MFCC features was also found satisfactory as compared to other features.
- Concluding any relation for speaker identification performance in each language is difficult since the distribution of speakers in different languages is non-uniform [10].

Table 2: The cross lingual speaker identification performance using different features and GMM-UBM classifier.

Training/ Testing	Features	Gaussian mixtures				P_i
		16	32	64	128	
E/H	LPCC	60	68	70	74	74
	MFCC	64	70	74	78	78
	multitaper MFCC	66	76	80	78	78
	τ^{th} - order MFCC	68	76	80	78	80
	IMFCC	60	64	70	68	70
	LFCC	64	70	70	76	76
	RFCC	62	70	72	78	78
	APGDF	60	66	70	76	76
	SCFC	70	78	80	78	80
	SSFC	58	64	72	74	74
H/E	LPCC	54	66	70	76	76
	MFCC	56	64	70	74	74
	multitaper MFCC	62	66	70	76	76
	τ^{th} - order MFCC	60	70	76	74	76
	IMFCC	62	64	70	68	70
	LFCC	64	68	74	76	76
	RFCC	58	76	74	70	74
	APGDF	62	64	70	68	70
	SCFC	66	70	78	74	78
	SSFC	60	68	74	72	74

V. CONCLUSION AND FUTURE WORK

In this work, an extensive study with different feature extraction techniques for monolingual, crosslingual and multilingual speaker identification was performed. Results indicated that the SCFC, LFCC, and multitaper MFCC were useful. Future work shall include exploring the combination of features for speaker identification for language mismatch condition.

Table 3: The multilingual speaker identification performance using different features and GMM-UBM classifier.

Training/ Testing	Features	Gaussian mixtures				P_i
		16	32	64	128	
E/Multi	LPCC	64	70	80	76	80
	MFCC	60	68	74	78	78
	multitaper MFCC	70	74	78	80	80
	τ^{th} - order MFCC	70	76	82	80	82
	IMFCC	64	68	74	76	76
	LFCC	64	70	80	82	82
	RFCC	68	74	80	80	80
	APGDF	66	66	74	76	76
	SCFC	68	74	80	82	82
	SSFC	68	74	78	78	78
H/Multi	LPCC	70	80	82	80	82
	MFCC	66	70	78	76	78
	multitaper MFCC	64	66	76	78	78
	τ^{th} - order MFCC	64	70	76	80	80
	IMFCC	60	64	70	74	74
	LFCC	70	70	78	80	80
	RFCC	58	68	74	70	74
	APGDF	62	64	70	72	72
	SCFC	64	72	80	76	80
	SSFC	58	66	78	78	78

REFERENCES

1. J. P. Campbell, Speaker recognition: A tutorial, Proc. IEEE, vol. 85, no. 9, pp. 1437-1462, 1997.
2. T. Kinnunen and H. Li, An Overview of text-independent speaker recognition?: From features to super vectors, Speech Commun., vol. 1, no. 1, pp. 12-40, 2009.
3. 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.
4. B. G. Nagaraja, Multilingual speaker identification, Ph.D. dissertation, Visvesvaraya Technological University, Belagavi, India, Oct. 2014.
5. H. Lei and E. Lopez, Mel, linear, and antimel frequency cepstral coefficients in broad phonetic regions for telephone speaker recognition, Interspeech, pp. 2323-2326, 2009.
6. Y. Wang and B. Lawlor, Speaker recognition based on MFCC and BP neural networks, 28th Irish Signals Syst. Conf., pp. 1-4, 2017.
7. N. Wang and L. Wang, Robust speaker recognition based on multi-stream features, IEEE International Conference on Consumer Electronics-China (ICCE-China), pp. 1-4, 2016.
8. X. Shi, H. Yang, and P. Zhou, Robust speaker recognition based on improved GFCC, 2nd IEEE International Conference on Computer and Communications, no. 4, pp. 1927-1931, 2016.
9. U. Bhattacharjee and K. Sarmah, A multilingual speech database for speaker recognition, IEEE Int. Conf. Signal Process. Comput. Control. ISPCC 2012, pp. 1-5, 2012.
10. S. Sarkar, K. S. Rao, D. Nandi, and S. B. S. Kumar, Multilingual speaker recognition on Indian languages, Proc. Annu. IEEE India Conf. INDICON 2013, pp. 1-5, 2013.
11. H. S. Jayanna, Limited data speaker recognition Ph.D. dissertation, Indian Institute of Technology Guwahati, Dept. of of Electronics & Communication Engg., Mar. 2009.
12. S. Chakroborty, A. Roy, and G. Saha, Improved closed set text-independent speaker identification by combining MFCC with evidence from flipped filter banks, Int. J. Signal Process., vol. 4, no. 2, pp. 114-121, 2007.
13. T. Kinnunen, R. Saeidi, F. Sedlak, K.A. Lee, J. Sandberg, M. Hansson-Sandsten and H. Li, Low-variance multitaper MFCC features: A case study in robust speaker verification, IEEE Trans. Audio, Speech Lang. Process., vol. 20, no. 7, pp. 1990-2001, 2012.
14. M. J. Alam, T. Kinnunen, P. Kenny, P. Ouellet, and D. O'Shaughnessy, Multitaper MFCC and PLP features for speaker verification using i-vectors, Speech Commun., vol. 55, no. 2, pp. 237-251, 2013.
15. T. Kinnunen, R. Saeidi, J. Sandberg, and M. Hansson-sandsten, What else is new than the Hamming window? Robust MFCCs for speaker recognition via multitapering, Proc. Interspeech, vol. 20, no. 7, pp. 2734-2737, 2010.
16. D. J. Thomson, Spectrum estimation and harmonic analysis, Proc. IEEE, vol. 70, no. 9, pp. 1055-1096, 1982.
17. M. Hansson and G. Salomonsson, A multiple window method for estimation of peaked spectra, IEEE Trans. Signal Process., vol. 45, no. 3, pp. 778{781, 1997.
18. K. S. Riedel and A. Sidorenko, Minimum bias multiple taper spectral estimation, IEEE Trans. Signal Process., vol. 43, no. 1, pp. 188-195, 1995.
19. Md. Sahidullah and G. Saha, A novel windowing technique for efficient computation of MFCC for speaker recognition, IEEE signal processing letters, vol. 20, no. 2, pp. 149-152, 2013.
20. X. Zhou, D. Garcia-Romero, R. Duraiswami, C. Espy-Wilson, and S. Shamma, Linear versus mel frequency cepstral coefficients for speaker recognition, Proc. IEEE Work. Autom. Speech Recognit. Understanding, ASRU 2011, pp. 559-564, 2011.
21. Md. Sahidullah, TomiKinnunen, and CemalHanili, A comparison of features for synthetic speech detection, Proc. Interspeech, pp. 2087-2091, 2015.
22. T. Hasan, S. O. Sadjadi, G. Liu, N. Shokouhi, H. Boril, and J. H. L. Hansen, CRSS systems for 2012 NIST speaker recognition evaluation, Proc. ICASSP, IEEE Int. Conf. Acoust. Speech Signal Process., pp. 6783-6787, 2013.
23. P. Rajan, T. Kinnunen, C. Hanilci, J. Pohjalainen, and P. Alku, Using group delay functions from all-pole models for speaker recognition, Proc. Annu. Conf. Int. Speech Commun. Assoc. INTERSPEECH, pp. 2489-2493, 2013.

Published By:
Blue Eyes Intelligence Engineering
and Sciences Publication (BEIESP)

24. T. Thiruvaran, E. Ambikairajah, and J. Epps, Group delay features for speaker recognition, Proc. 6th Int. Conf. Information, Commun. Signal Process. ICICS, vol. 2, no. 2, pp. 1-5, 2007.
25. K. KuldipPaliwal, Spectral subband centroid features for speech recognition, Proc. IEEE International Conference on Acoustics, Speech and Signal Processing, vol. 2, pp. 617-620. 1998.
26. N. P. H. Thian, C. Sanderson, and S. Bengio, Spectral subband centroids as complementary features for speaker authentication, Proc. Biometric Authentication, pp. 631-639, 2004.
27. J. Kua, Investigation of spectral centroid magnitude and frequency for speaker recognition, Proc. Odyssey, pp. 34-39, 2010.
28. E. Scheirer and M. Slaney, Construction and evaluation of a robust multi feature speech/music discriminator, Proc. IEEE Int. Conf. Acoust. Speech, Signal Process., vol. 2, pp. 2-5, 1997.
29. B. C. Haris, G. Pradhan, A. Misra, S. Shukla, R. Sinha, and S. R. M. Prasanna, Multi-variability speech database for robust speaker recognition, Proc. Natl. Conf. Commun., pp. 1-5, 2011.
30. G. Pradhan and S. R. M. Prasanna, Significance of vowel onset point information for speaker verification," Int. J. Comput. Comm. Tech., vol. 2, no. 6, pp. 56-61, 2011.
31. J. P. Campbell Jr, Testing with the YOHO CD-ROM voice verification corpus, Proc. Int. Conf. Acoust. Speech, Signal Process. ICASSP-95, vol. 1, pp. 341-344, 1995.
32. Geoffrey Durou, Multilingual text-independent speaker identification, Proc. MIST'99 Workshop, pp. 115-118, 1999.

Published By:
Blue Eyes Intelligence Engineering and Sciences Publication (BEIESP)