


Noise Robust Speaker Identification Using RASTA–MFCC Feature with Quadrilateral Filter Bank Structure

S. Selva Nidhyananthan¹ · R. Shantha Selva Kumari¹ ·
T. Senthur Selvi¹ 

Published online: 8 August 2016
© Springer Science+Business Media New York 2016

Abstract This paper motivates the use of Relative Spectra–Mel Frequency Cepstral Coefficients (RASTA–MFCC) feature extracted from the newly designed Quadrilateral filter bank structure and Gaussian Mixture Model–Universal Background Model (GMM–UBM) for improved text independent speaker identification under noisy environment. Unlike neural network model which requires retraining of entire database when a new sample is added to it, GMM–UBM model does not require retraining of entire database which leads to easier and faster processing. RASTA–MFCC is found to be more robust to noisy environment compared with traditional MFCC method. MFCC is an efficient feature for identifying the speaker as it has speaker specific information capturing ability. RASTA processing of speech improves the performance of recognizer in the presence of convolution and additive noise. This work combines the better of these two processes to yield RASTA–MFCC feature which is robust to noise and also proposes a new Quadrilateral filter bank structure which approximates the response of cochlear membrane of human ear to effectively capture the feature vectors. The proposed Quadrilateral filter bank structure with RASTA–MFCC feature and GMM–UBM modeling for speaker identification demonstrates supremacy over triangular and Gaussian filter banks and offers a speaker identification accuracy of 97.67 % for the MEPCO noisy speech database with 50 speakers.

Keywords Cepstral mean normalization (CMN) · Equivalent rectangular bandwidth (ERB) · Gaussian Mixture Model–Universal Background Model (GMM–UBM) · Mel Frequency Cepstral Coefficients (MFCC) · Relative Spectra processing (RASTA)

✉ T. Senthur Selvi
selvi.sen91@gmail.com

S. Selva Nidhyananthan
nidhyan@mepcoeng.ac.in

R. Shantha Selva Kumari
rshantha@mepcoeng.ac.in

¹ Department of ECE, Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, India

1 Introduction

The prime motto of speech analysis is efficiently characterizing the information present in the speech signal either to identify the speech or to identify the speaker. Since both speech recognition and speaker identification involve pattern recognition, the speech analysis techniques are almost similar for both. In speech signal analysis, the kind of information to be retained in the form of feature vector depends on the application for which the speech signal is analyzed. For example, speaker-dependent attributes are obviously fully relevant for speaker identification, but those attributes are often superfluous for speech recognition. In speaker identification N comparisons are required between the test pattern and the stored N enrolled patterns. The speaker is identified among N speakers in the database based on minimum absolute probability of error.

The speaker can be identified relevant to the text spoken or irrelevant to the text spoken. The former task is called as text-dependent speaker identification and the later is called as text-independent speaker identification. Text-dependent is the simplest of these two, wherein a small set of specific words is used in enrollment phase and the words from same set must be used by the speaker in the test phase in order for correct identification. Text-independent speaker identification system imposes no boundary or limitation on the words or phrases that can be used for identifying the speaker. Since the speaker is provided with the freedom of using any utterance during testing irrespective of the utterance used during enrollment, this mode of speaker identification is comparatively complex and challenging.

Efficient representation of speaker oriented information present in the speech signal is of at most important to have better text independent speaker identification system. Thus the speaker-dependent aspects of speech have to be represented economically with reduced dimensionality. The selected features must have large inter-speaker variability and small intra-speaker variability and must have robustness against mimicry. Different researchers have investigated the usage of features such as MFCC, BFCC, Perceptual Linear Prediction (PLP) coefficients [1], Linear Predictive Residual Cepstral Coefficients (LPRCC). Few researchers have used delta (Δ) and delta-delta ($\Delta-\Delta$) coefficients of the features to improve the identification accuracy. Most of the researchers agreed upon the supremacy of MFCC features over other features in speaker identification and speech recognition. GMM modeling with multivariate Gaussian distribution technique best configures the human vocal tract, by which the identity of the individual speaker is best reflected. Performance of GMM decays under noisy conditions [2]. Compared to GMM, the universal background model based GMM reduces the dimensionality and computational complexity. The Universal background model encompasses the overall characteristic of the population in a single pool [3, 4]; at the same time it adapts the pool to the individual speaker.

Challenging factors in accurate speaker identification system are noise due to hostile environment and channel distortions [5]. When speech recordings are done using microphones with different sensitivities, the channel effects become prominent. Variations in speech spectral component due to noise can be handled effectively by RASTA processing [6]. This improves the performance of the identification system in the presence of convolution and additive noise [7]. Cepstral mean normalization minimizes the degradation in perceived quality of speech by channel equalization [8].

In the proposed work, the combined RASTA-MFCC feature for improving the performance of the identification system under noisy environment is analyzed for different filter bank structures. It is evident from the statistical result that the performance of

RASTA–MFCC feature surpasses the conventional MFCC feature in unknown channel and noisy environment.

The organization of the paper is as follows: Overall speaker identification system and preprocessing are discussed in Sect. 2. The new RASTA–MFCC feature extraction process is discussed in Sect. 3 with detailed explanation. Section 4 describes the design aspects of the newly designed quadrilateral filter bank structure. In Sect. 5, modeling of features via GMM–UBM techniques is dealt and the obtained results are analyzed. Finally, Sect. 6 concludes the proposed work.

2 Speaker Identification Systems

Speaker identification can be carried out in two stages as shown in Fig. 1. Pattern representation of the speech samples followed by modeling of patterns/vectors is done in enrollment/training stage. In the testing stage, the log likelihood ratio of the test speaker model is one-to-one compared with all the stored models to find the minimum probability of error.

For dimensionality reduction, the voiced and unvoiced regions are separated. Since the entire speaker specific information is present in the voiced region [9], it is retained for further processing. Removal of unvoiced and silence regions from speech samples are done using energy based thresholding technique which reduces the computational requirements. To make use of intermittent nature of the speech signal, the voiced regions of the speech is segmented into frames and each frame is windowed to provide smooth tapering.

Even though speech signal is quasi-periodic, when it is processed in segmented frames of 10–30 ms duration, the characteristics of speech resembles the characteristics of stationary and periodic signals, mainly at the occurrences of vowels. Biological production of speech is merely a filtering operation in which voiced sound is produced by periodic source exciting a vocal tract filter. Over the duration of a frame the speech is interpreted as the stationary signal because of the tendency of the signal to gradually change its characteristics between sounds. If framing is done with non-overlapping between frames then there may be loss of information due to the transition between adjacent frames. Usually overlap

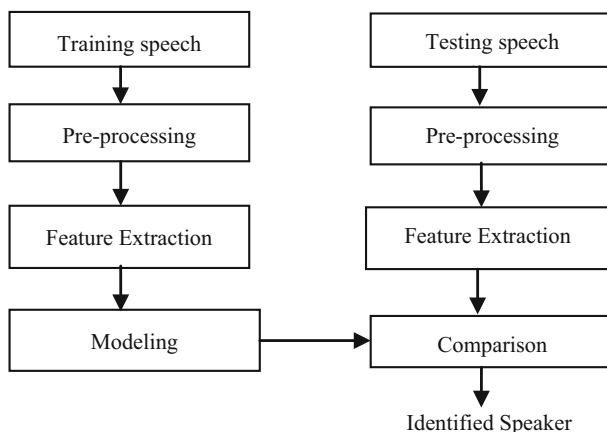


Fig. 1 Speaker identification system

size of more than 50 % of frame size yields better result [10]. Researches on speech and speaker recognition unanimously agree that Hamming window best suits for speech processing applications. The window length is kept same as the individual frame length. The choice of window shape for producing desired smoothing [11] depends on its effect in speech analysis. Windowing is done on each frame in order to taper the signal to zero at the beginning and the end of the frame.

3 Noise Robust RASTA–MFCC

RASTA filtering is applied on the windowed speech signal to minimize the noise effects in the speech signal, especially convolution and additive noise effects [12]. Filtering is followed by the extraction of MFCC from the RASTA filtered signal in order to yield RASTA–MFCC features. The steps followed in obtaining RASTA–MFCC feature is depicted in Fig. 2.

RASTA processing improves the performance of a recognizer in noisy conditions. RASTA processing compensates the effect of abrupt spectral change in speech signal by means of filtering. Fast spectral changes in Consecutive frames are alleviated by low pass filtering [13] through smoothing process. In general, bigger the auditory structures, more the sensitivity to lower speech/sound frequencies. In the mammal family, humans have relatively less sensitivity to lower frequency sounds. RASTA processing involves computation of power spectrum of critical band, filtering the time trajectory of compressed spectral component, static nonlinear transformation followed by multiplication with equal loudness curves. Finally computes all-pole model of the spectrum.

Lower cutoff frequency of the filter determines the fastest spectral change whereas the higher cutoff frequency determines the preserved spectral change. Computation of squared magnitude of FFT follows RASTA filtering. Pre-emphasis emphasizes the energy of the high frequency contents of the squared magnitude spectrum. The pre-emphasis that equalizes the speech spectral tilt is given in Eq. (1) with the pre-emphasis factor α value 0.97.

$$\hat{s}(n) = s(n) - \alpha s(n - 1) \tag{1}$$

$s(n)$ is the n th instant of the speech signal, $s(n - 1)$ is the $n - 1$ th instant of the speech signal, $\hat{s}(n)$ is the n th instant of the pre-emphasized signal.

Human auditory perception is a nonlinear process. Mel scale mapping from linear frequency resembles human auditory pattern. As shown in Fig. 3 Mel scale mapping is

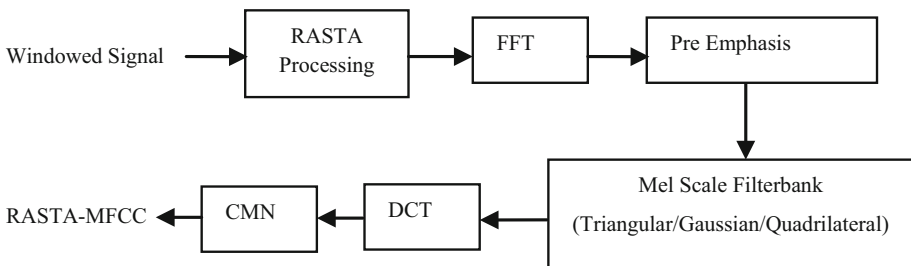


Fig. 2 RASTA–MFCC feature extraction process

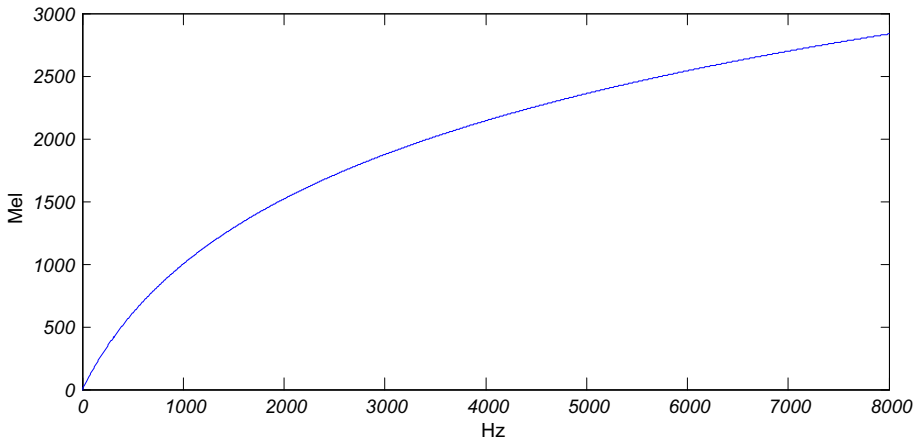


Fig. 3 Relation between Mel and linear frequency

approximately linear for frequencies up to 1 kHz and logarithmic afterwards. The relation between Hertz and Mel scale [14] is given as follows.

$$mel(f) = 2595 \log_{10} \left(1 + \left(\frac{f_{\text{Hz}}}{700} \right) \right) \tag{2}$$

Conventionally, the critical band triangular shaped filters are residing on the Nyquist range. The transforms of the filters are made symmetrical about the Nyquist frequency. As shown in Fig. 4, the Mel axis filter bank is constructed with 40 non uniform filters. In order to have smooth transition between adjacent critical bands and to preserve the correlation among them Gaussian filter bank is also developed with 40 non uniform filters as shown in Fig. 5.

After Mel scale warping, Mel spectral coefficients are obtained, for which discrete cosine transform is taken in order to yield the Rasta Mel frequency Cepstral coefficients. MFCC extraction is similar to cepstrum calculation except the Mel scale frequency axis. By applying DCT reduced data set representation is obtained. Equation (3) gives RASTA–MFCC coefficients.

$$c(l) = \frac{\sqrt{2}}{M} \sum_{m=1}^M X'(m) \cos \left(\frac{l\pi}{M} \left(m - \frac{1}{2} \right) \right) \tag{3}$$

where $X'(m)$ are the Mel spectral coefficients, M is the number of filters.

To help in minimizing the effect of channel in speech recording noise spectral subtraction [15] and Cepstral Mean Normalization (CMN) methods can be used. Since the former method has the problem of estimating the noise [16], the later method is preferred to mitigate the effect of variable communication environment. In CMN, the average value of the RASTA–MFCC coefficients over the whole length of the speech is subtracted from each frame as follows.

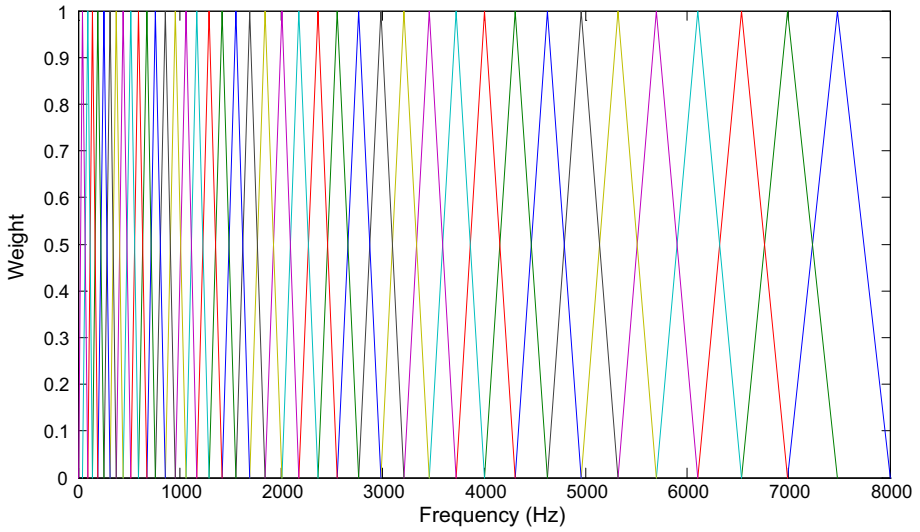


Fig. 4 Triangular filter bank structure

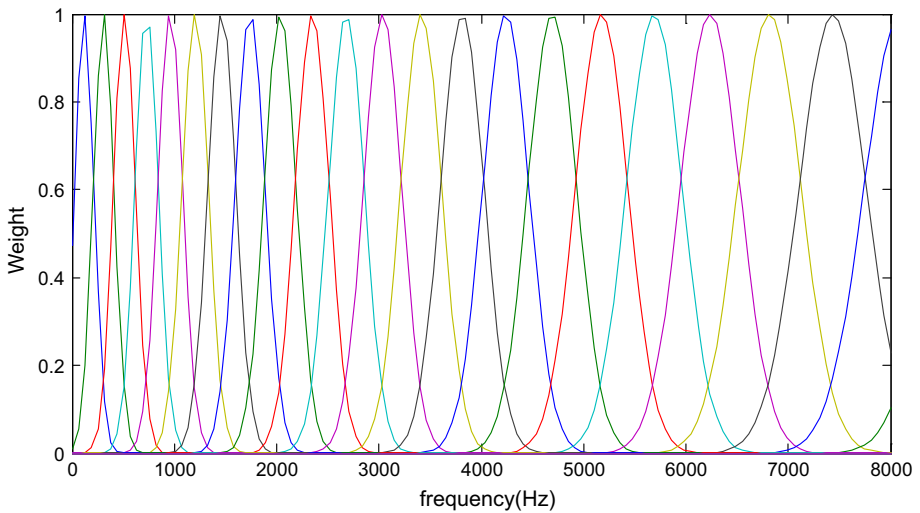


Fig. 5 Gaussian filter bank structure

$$c_i = c_i - \frac{1}{N} \sum_{k=1}^N c_{ik} \tag{4}$$

where, c_{ik} is the i th feature element in the k th frame. The resultant feature after Cepstral mean normalization, a post processing step, yields the noise robust RASTA-MFCC Feature.

4 Quadrilateral Filter Bank Structure

In the triangular shaped or Gaussian shaped Mel filter bank design, the speaker dependent information around the lower frequency range of each filter bin of the filter bank is not given much importance to encompass as much energy as possible. Hence, a new filter bank with quadrilateral shaped filter bins is designed in which the lower frequency of the current filter bin is the first intermediate frequency of the previous filter bin. The First and the last filter bin’s center frequency are determined from Moore and Grasberg’s ERB (Equivalent Rectangular Bandwidth) expression as given by [17].

$$af_{c_i}^2 + bf_{c_i} + c = \frac{1}{2} (f_{high_i} - f_{low_i}) \tag{5}$$

$$ERB_i = 24.7(0.00437f_{c_i} + 1) \tag{6}$$

For the remaining filter bins the center frequency is found using Eq. (7),

$$f_{c_i} = f_{low_i} + (i - 1) \left[\frac{f_{high_i} - f_{low_i}}{N + 2} \right] \tag{7}$$

The lower and upper frequency of each filter bin is found using the following Eqs. (8) and (9),

$$(700 + f_{c_i})^2 = (700 + f_{low_i} + 2ERB_i)(700 + f_{low_i}) \tag{8}$$

$$f_{high_i} = f_{low_i} + 2ERB_i \tag{9}$$

The two intermediate frequencies f_{int1} and f_{int2} are found using the Eqs. (10) and (11),

$$f_{int1} = 0.25 * (f_{high_i} + f_{low_i}) \tag{10}$$

$$f_{int2} = 0.75 * (f_{high_i} + f_{low_i}) \tag{11}$$

where, $a = 6.23 \times 10^{-6}$; $b = 93.39 \times 10^{-3}$; $c = 28.52$, f_{c_i} is the i th center frequency of the filter bin, f_{high_i} is the upper frequency range of the i th filter bin, f_{low_i} is the lower frequency range of the i th filter bin, f_{int1} is the first intermediate frequency of the i th filter bin, f_{int2} is the second intermediate frequency of the i th filter bin, $i = 1, 2 \dots N$, N is the total number of filter bins in the filter bank.

The amplitude of the four vertices in each quadrilateral bin is [0, 0.7, 1, and 0]. The value 0.7 is found to be optimum to height of the second vertices after a series of test for the values in between the range [0.5, 1.0]. The designed Quadrilateral filter bank structure is placed in Mel frequency scaling in order to closely approximate the human cochlear membrane and the resultant filter bank structure is shown in Fig. 6.

5 Modeling and Result Analysis

The objective of GMM–UBM modeling is estimating the test model parameters that match with the distribution of the training feature vector. UBM model is trained by computing λ_p which is constituted by mean vector, variance vector and weight vector. Background model first takes the common characteristics of the population then adjust it to the individual. The

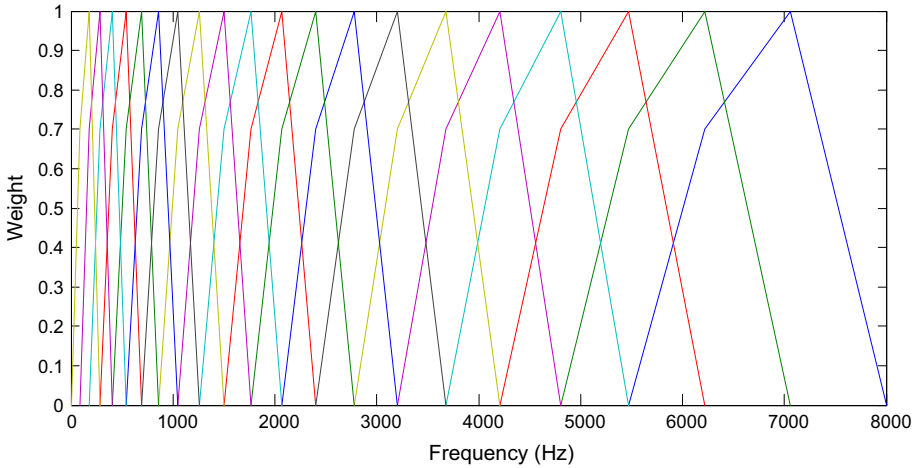


Fig. 6 Quadrilateral filter bank structure

Log Likelihood Ratio (LLR) score is the tool to identify the speaker under test. The test speaker is compared with the enrolled speakers in terms of their likelihood and the one match with maximum LLR score is declared as the identified speaker. In GMM-UBM, the background model is taken into consideration. Speaker model is represented using background model and adapted model. The density function is calculated for GMM-UBM with 256 mixtures.

$$P_r(i/x_t, \lambda) = \frac{w_i p_i(x_t)}{\sum_{j=1}^M w_j p_j(x_t)} \tag{12}$$

In Expectation Maximization algorithm, the values of the model parameters change for every iteration. New coefficients are calculated using the Eqs. (13), (14) and (15) at every iteration.

$$\mu_i^{new} = \alpha_i^m E(x)_i + (1 - \alpha_i^m) \mu_i \tag{13}$$

$$\sigma_i^{new} = \alpha_i^y E(x^2)_i + (1 - \alpha_i^y) (\sigma_i^2 + \mu_i) - \mu_i^{new} \tag{14}$$

$$w_i^{new} = [\alpha_i^w n_i / T + (1 - \alpha_i^{new}) w_i] \gamma \tag{15}$$

where $\alpha = n(i)/(n(i) + r)$.

GMM distribution represents the best distribution of feature vectors for hypothesis $H0$. The UBM is used for modeling the alternative hypothesis $H1$ in the likelihood ratio test. For a given set of N background speaker models, the alternative hypothesis $H1$ is represented by Eq. (16),

$$p(X/\lambda'_p) = \max(p(X/\lambda_1)p(X/\lambda_2) \dots p(X/\lambda_N)) \tag{16}$$

With UBM treated as prior model, a speaker specific model is derived by using maximum likelihood estimation. For a given T independent and identically distributed observations, $X = \{x_1, x_2, x_3, x_4 \dots x_T\}$, the joint likelihood ratio is determined using Eq. (17).

$$E(LLR(x)) = \frac{1}{T} \sum_{t=1}^T (\log p(x_t | \lambda_{target}) - \log p(x_t | \lambda_{ubm})) \quad (17)$$

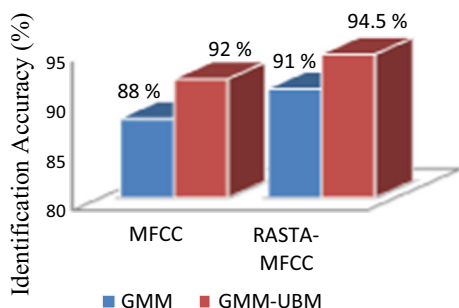
MEPCO speech biometric database is used in the proposed speaker identification system with 50 speakers and among them 10 speakers additionally perform the task of imposters. The recording process for MEPCO speech biometric database was done using Gold Wave version 5.58 software and Condenser microphone at 16 kHz sampling rate mono mode recording with PCM coding. In order to accommodate this time varying nature of speech, recording are done in different days. In order to test the robustness of the proposed speaker identification system against real world noise, the recordings are done in classroom environment having disturbances like other student's speech, humming noise from Air conditioner, ceiling fan noise and electricity generator noise. 6 speech samples are recorded from each speaker and each recording spans 3 s duration.

Since only voiced speech has useful speaker-specific information, the unvoiced and silence regions of speech waves are removed using energy based thresholding technique. Almost half the processing requirement is reduced after silence removal. Since the speech signal is assured to be short time stationary, voiced speech is divided into overlapping frames of length 256 samples with amount of overlapping 50 %. Compared to other windows, Hamming window produces much less spectral leakage. Hence framing is followed by hamming window process. Pre-emphasis is done with a Pre-emphasis factor of 0.97. In order to have speech features vigorous against noise, RASTA filtering is performed on the windowed speech frames. To obtain RASTA-MFCC feature, 40 filters filter bank is implemented for both triangular and Gaussian filter banks. The feature models are obtained by having 256 mixtures in GMM-UBM models. Out of the 6 sessions of speech recording for every speaker, first three sessions are used for training the speaker model, rest three sessions are used for testing the identity of the speaker. For every speaker the ratio of number of correctly identified session to the total number of sessions is calculated. This ratio in percentage is treated as the identification accuracy of that particular speaker. Similarly, the identification accuracy of all the 50 speakers is calculated. The average of all these 50 identification accuracies is the identification accuracy of the proposed speaker identification system.

$$\text{Identification Accuracy (\%)} = \frac{\text{Number of sessions correctly identified}}{\text{Total number of sessions}} \times 100 \quad (18)$$

The proposed speaker identification system provides an efficient identification of 94.5 % for triangular filter bank design and 96 % for Gaussian filter bank design. The

Fig. 7 Comparison between different features and modeling for triangular filter bank structure



performance of RASTA–MFCC feature is compared with traditional MFCC feature in text independent speaker identification system under noisy environment. It is found that the RASTA–MFCC feature is more robust and provides an identification accuracy of 97.67 % in the case of Quadrilateral filter bank with the speech database size of 50 speakers while the MFCC method provides an accuracy of 88 %. GMM–UBM modeling is used for its effective resistance towards imposter attack. A bar chart for the performance comparison of different GMM modeling methods with different MFCC features is shown in Figs. 7, 8. A comparison between the performances of various filter bank structures with RASTA–MFCC features modeled using GMM–UBM is shown in Fig. 9.

Fig. 8 Comparison between different features and modeling for gaussian filter bank structure

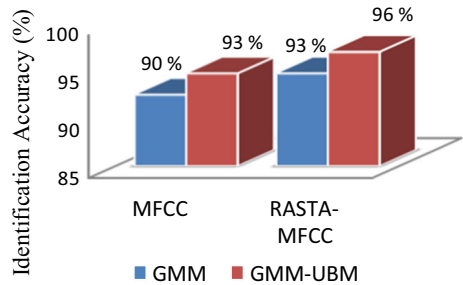


Fig. 9 Comparison between different filter bank structures

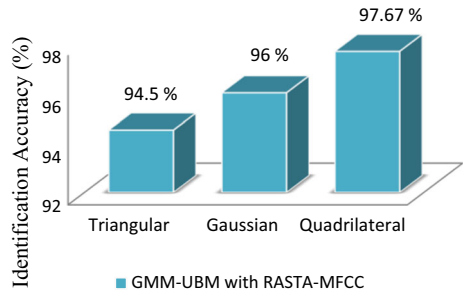


Table 1 Comparison of various speaker identification works

S. No.	Speaker identification works	Feature	Database size	Accuracy (%)
1	Marco et al. [18]	RASTA-PLP	24 speakers	80
2	Reynolds [19] (noise robust)	MFCC	49 speakers telephone speech	80.8
3	Revathi et al. [20]	PLP	50 speakers from TIMIT	91
4	Gomez [21]	Multiple parametric self-organizing maps	40 speakers from CSLU	93
5	Triangular filter bank	RASTA–MFCC	50 speakers	94.5
6	Gaussian filter bank	RASTA–MFCC	50 speakers	96
7	Proposed quadrilateral filter bank	RASTA–MFCC	50 speakers	97.67

All the 10 imposter speakers have been correctly identified. Table 1 show that the proposed Quadrilateral filter bank with RASTA–MFCC feature outperforms other speaker identification techniques. The reason behind this is that the proposed filter bank encompasses more low frequency, high energy speaker specific information than high frequency information for speaker modeling.

6 Conclusion

In this paper, text independent speaker identification under noisy environment is implemented using RASTA–MFCC as feature vector and GMM–UBM as the modeling. A new Quadrilateral filter bank structure is designed and its performance is found to be better than conventional filter banks. Experimental results show that the RASTA–MFCC features with Quadrilateral filter banks are more robust to noisy environment than triangular and Gaussian filter banks. The UBM adaptation is faster than GMM training. The quality of UBM is better than GMM when small training segments on the order of 2–5 s. Only the detection time of UBM is longer than GMM. Speaker identification system may have applications in banking over telephone, attendance systems, computer security, database access systems, and forensics.

References

1. Atal, B. S. (1974). Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification. *The Journal of the Acoustical Society of America*, 55(6), 1304–1312. doi:[10.1121/1.1914702](https://doi.org/10.1121/1.1914702).
2. Reynolds, D.A. (2008). *Gaussian mixture models*. Lexington, MA: MIT Lincoln Laboratory.
3. Reynolds, D. A., Quatieri, T. F., & Dunn, R. B. (2013). Speaker verification using adapted gaussian mixture models. *IEEE Transactions on Audio, Speech and Language Processing*. doi:[10.1006/dspr.1999.0361](https://doi.org/10.1006/dspr.1999.0361).
4. Bhattacharjee, U., & Sarmah, K. (2012). GMM–UBM based speaker verification in multilingual environments. *IJCSI International Journal of Computer Science Issues*, 9(6), 2.
5. Xiaojia, Z., Yang, S., & De Liang, W. (2011). Robust speaker identification using auditory features and computational auditory scene analysis. *IEEE Proceedings of the ICASSP-2008*. doi:[10.1109/ICASSP.2008.4517928](https://doi.org/10.1109/ICASSP.2008.4517928).
6. Hermansky, H., Morgan, N., Bayya, A., & Kohn, P. (1992). RASTA-PLP speech analysis technique. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1, 121–124. doi:[10.1109/ICASSP.1992.225957](https://doi.org/10.1109/ICASSP.1992.225957).
7. Skowronski, M. D., & Harris, J. G. (2003). Improving the filter bank of a classic speech feature extraction algorithm. *IEEE International Symposium on Circuits and Systems*. doi:[10.1109/ISCAS.2003.1205828](https://doi.org/10.1109/ISCAS.2003.1205828).
8. Schwarz, R., et al. (1994). Comparative experiments on large vocabulary speech recognition. *IEEE International Conference on Acoustics, Speech, and Signal Processing*, 1(1), 561–564. doi:[10.1109/ICASSP.1994.389232](https://doi.org/10.1109/ICASSP.1994.389232).
9. Un, C., & Lee, H. (1980). Voiced/unvoiced/silence discrimination of speech by delta modulation. *IEEE Transaction on Acoustics, Speech and Signal Processing*, 28(4), 398–407. doi:[10.1109/TASSP.1980.1163424](https://doi.org/10.1109/TASSP.1980.1163424).
10. Picone, J. (1993). Signal modeling techniques in speech recognition. *Proceedings of the IEEE*, 81(9), 1215–1247. doi:[10.1109/5.237532](https://doi.org/10.1109/5.237532).
11. Deller, J. R., Proakis, J. G., & Hansen, J. H. L. (1993). *Discrete time processing of speech signals*. London: Macmillan.
12. Hermansky, H., & Morgan, N. (1994). RASTA processing of speech. *IEEE Transaction Speech Audio Process*, 2(4), 578–589. doi:[10.1109/89.326616](https://doi.org/10.1109/89.326616).

13. Gaubitch, N. D., Brookes, M., & Naylor, P. A. (2013). Blind channel magnitude response estimation in speech using spectrum classification. *IEEE Transactions on Audio, Speech and Language Processing*. doi:[10.1109/TASL.2013.2270406](https://doi.org/10.1109/TASL.2013.2270406).
14. Togneri, R., & Pullela, D. (2011). An overview of speaker identification and accuracy. *IEEE Circuits and Systems Magazine*. doi:[10.1109/MCAS.2011.941079](https://doi.org/10.1109/MCAS.2011.941079).
15. Stockham, T., Cannon, T., & Ingebretsen, R. (1975). Blind deconvolution through digital signal processing. *Proceedings of the IEEE*, 63, 678–692. doi:[10.1109/PROC.1975.9800](https://doi.org/10.1109/PROC.1975.9800).
16. Wojcicki, K., & Loizou, P. (2012). Channel selection in the modulation domain for improved speech intelligibility in noise. *Journal of the Acoustical Society of America*, 131(4), 2904–2913. doi:[10.1121/1.3688488](https://doi.org/10.1121/1.3688488).
17. Moore, B. C. J., & Glasberg, B. R. (1983). Suggested formulae for calculating auditory-filter bandwidths and excitation patterns. *The Journal of the Acoustical Society of America*, 74(3), 750–753.
18. Grimaldi, M., & Cummins, F. (2008). Speaker identification using instantaneous frequencies. *IEEE Transactions on Audio, Speech and Language Processing*. doi:[10.1109/TASL.2008.2001109](https://doi.org/10.1109/TASL.2008.2001109).
19. Reynolds, D. A., & Rose, R. C. (1995). Robust text-independent speaker identification using gaussian mixture models. *IEEE Transaction on Speech Audio Processing*, 3(1), 72–83. doi:[10.1109/89.365379](https://doi.org/10.1109/89.365379).
20. Revathi, A., Ganapathy, R., & Venkataramani, Y. (2009). Text independent speaker recognition and speaker independent speech recognition using iterative clustering approach. *International Journal of Computer Science and Information Technology*, 1(2), 30–42.
21. Gomez, P. (2011). A text independent speaker recognition system using a novel parametric neural network. *Proceedings of International Journal of Signal Processing, Image Processing and Pattern Recognition*, 4(4), 1–16.



S. Selva Nidhyanathan received his B.E. degree in Electronics and Communication Engineering from Manonmanium Sundaranar University, Tirunelveli, in 1999, M.E. degree in Communication Systems from Anna University, Chennai, in 2005 and Ph.D. degree Anna University, Chennai, in 2014. He has 14 years of teaching experience and he is currently working as Associate Professor in the department of Electronics and Communication Engineering at Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu. His current research interest includes Signal Processing and Image Processing.



R. Shantha Selva Kumari received her B.E. degree in Electronics and Communication Engineering from Bharathiyar University, in 1987, M.S. degree in Electronics and Control from Birla Institute of Technology, Pilani, in 1994 and Ph.D. degree in Bio Signal Processing from Manonmanium Sundarana University, Tirunelveli, in 2008. She has 27 years of teaching experience and she is currently working as Senior Associate Professor and Head in the department of Electronics and Communication Engineering at Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu. Her current research interest includes Signal Processing, Wavelets and its Applications and Neural Networks.



T. Senthur Selvi received her B.E. degree in Electronics and Communication Engineering from Anna University, Chennai in 2012, and completed M.E. (Communication Systems) at Mepco Schlenk Engineering College, Sivakasi, Tamil Nadu, in 2014. Her areas of interest includes Digital Speech Signal processing, Digital Image processing.