International Journal of Speech Technology (2018) 21:185–192
https://doi.org/10.1007/s10772-018-9494-9

Robust noise MKMFCC–SVM automatic speaker identification

Osama S. Faragallah[1,2]

Received: 7 February 2017 / Accepted: 22 January 2018 / Published online: 14 February 2018

Abstract
This paper proposes robust noise automatic speaker identification (ASI) scheme named MKMFCC–SVM. It based on the Multiple Kernel Weighted Mel Frequency Cepstral Coefficient (MKMFCC) and support vector machine (SVM). Firstly, the MKMFCC is employed for extracting features from degraded audio and it uses multiple kernels such as the exponential and tangential and for MFCC's weighting. Secondly, the extracted features are then categorized with the SVM classification technique. A comparative study is performed between the proposed MKMFCC–SVM and the MFCC–SVM ASI schemes using the MKMFCC and MFCCs with five schemes for extracting features from telephone-analogous and noisy-like degraded audio signals. Experimental tests prove that the proposed MKMFCC–SVM ASI scheme yields higher identification rate in noise presence or degradation.

Keywords Automatic speaker identification (ASI) · MKMFCCs · SVM

1 Introduction

Enhancing automatic speaker recognition (ASR) systems has become an attractive challenge due to the growing needs for secure access or criminalistics inquiries. The main objective of ASR is to determine and recognize speaker personality, regardless of what speaker is clarifying (Dharanipragada et al. 2007; Huang et al. 2016; Shuling 2009; Gandhiraj and Sathidevi 2007). The ASR includes both verifying and identifying phases. In automatic speaker verification (ASV), speaker's speech is matched according to his/her pattern within the database and categorized either customer or imposter (Furui 1981). ASV systems can be usually utilized in many security fields such as telephone transactions. With automatic speaker identification (ASI), speech talking of anonymous speaker is tested and matched with patterns of all recognized speakers to determine the top matched speaker (Xu and Yang 2016; Li and Gao 2016; Hossain et al. 2007). ASI can be divided into either closed or open sets. Closed set ASI include that speaker under test was previously recognized to be one from finite set of speakers. Open set ASI involves the preference of defining declaring that test speaker may not belong to any one from recognized speakers.

ASR includes two phases stages, named, feature extracting and classification phases. The feature extracting phase may be thought as data reducing procedure with the potential of capturing main speaker features with reduced data as possible. There exist several schemes for speech features extraction using various coefficients types like linear prediction coefficients (LPCs) (Mellahi and Hamdi 2015), linear prediction cepstral coefficients (LPCCs) (Polur and Miller 2005), Mel-frequency cepstral coefficients (MFCCs) (Selva Nidhyananthan et al. 2016), and multiple kernel weighted MFCCs (Subba Ramaiah and Rajeswara Rao 2016).

Classification phase may be thought as a procedure that includes two stages named as; speaker modeling/ matching stages. In speaker modeling stage, the speaker is registered in the system with extracted features resulted from training data. If data sample of anonymous speaker is received, feature matching schemes can be utilized for mapping features of input speech data sample to a pattern that may correspond to a recognized speaker. The combination of speaker modeling/matching schemes may be known as

✉ Osama S. Faragallah
osam_sal@yahoo.com; o.salah@tu.edu.sa

[1] Department of Computer Science and Engineering, Faculty of Electronic Engineering, Menoufia University, Menouf 32952, Egypt

[2] Department of Information Technology, College of Computers and Information Technology, Taif University, Al-Hawiya 21974, Kingdom of Saudi Arabia

classifier. Classification schemes employed in ASI systems may cover Gaussian mixture models (GMMs) (Ding and Yen 2015; Qian et al. 2008), vector quantization (VQ), hidden Markov models (HMMs) (Polur and Miller 2005), ANNs (Galushkin 2007; Hayati and shirvany 2007) and SVM (Naeeni et al. 2010; Boujelbene et al. 2010; Zergat and Amrouche 2014).

In this paper, an efficient robust noise MKMFCC–SVM method for ASI is presented. The proposed method utilizes the MKMFCC as feature parameterization with multiple kernels such as the exponential and tangential to weight the MFCC's and the SVM for classification. The cepstral features combining the Mel filter bank tangential/exponential functions were utilized in cepstral coefficient parameterization. Multiple kernel weighted functions may help in considering low/ high energy frames of recognized audio signal, such that no frames dropped out. The paper remainder may be arranged as follow. Section 2 explores feature extraction using the MKMFCC. Then, the SVM is described in Sect. 3. Section 4 detailed the proposed MKMFCC–SVM ASI. Section 5 presents the utilized data sets and test results. And finally, Sect. 6 concludes the paper.

2 MKMFCC feature extraction

The MKMFCC employs two distinct kernel functions for the MFCC coefficients weighting (Ramaiah and Rao 2016). The kernel weighting offers a natural way for mixing and integrating various data types. Also, flexible mixture of suited kernel design and modern kernel schemes proved the superiority of such class of methods whose statistical and computational characteristics are well known by several machine learning methods. The MKMFCC is illustrated in Fig. 1 and detailed steps are given as follows.

2.1 Pre-emphasis

The pre-emphasis stage is employed for flattening speech spectrum, as it increases the high frequency band amplitude and reduce the low frequency band. It can be estimated by,

$$B(m) = A(m) - C \star A(m-1) \tag{1}$$

where *C, A, B, m* are constant value, input signal, output signal, and speech signal samples.

2.2 Framing

The speech signal sample is split into short L blocks of M samples. The speech block length is ranged as 20–40 ms. The neighbouring blocks are unattached by R factor; where $R < M$.

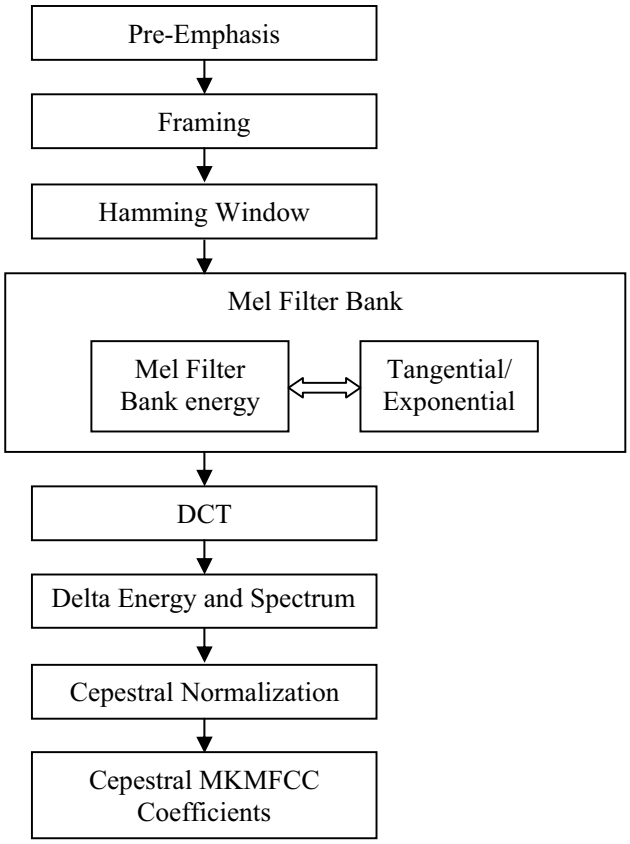

Fig. 1 Block diagram of MKMFCC (Ramaiah and Rao 2016)

2.3 Hamming windowing

During hamming window stage, all close frequencies in speech streams are integrated together. The hamming windowing can be represented as $W(m) : 1 \le m \le M-1$. The speech signal after employing windowing can be computed as,

$$B(m) = A(m) \star W(m) \tag{2}$$

where $W(m)$ is the hamming window and it is computed as,

$$W(m) = 0.56 - 0.46\left(\frac{2\pi m}{M-1}\right); 0 \le m \le M-1. \tag{3}$$

2.4 Fast fourier transform (FFT)

During FFT stage, the speech signal are FFT transformed. The block power spectrum can be computed as,

$$P_l(k) = \frac{1}{M}\left|A_l(k)\right|^2 \tag{4}$$

The Discrete Fourier Transform (DFT) of correspondent block can be estimated as,

$$A_l(k) = \sum_{m=1}^{M} B(m) \cdot e^{-j2\pi km}; 1 \le k \le K \tag{5}$$

where k is the DFT length and $B(m)$ covering M sample long analysis window.

2.5 Mel filter bank processing

Signal frequencies will be filtered using triangular filter for estimating filter spectral components weighted sum and Mel scale triangular filter output border. Figure 2 illustrates the Mel scale filter bank.

The high and low F_H/F_L frequency spectral components of periodogram estimations must be considered. The filter locations have equivalent space in Mel frequency.

$$MEL(f) = 1125 \times \ln\left(1 + \frac{f}{700}\right) \tag{6}$$

The Mel Filter bank can be estimated with FFT as

$$G(l) = (mFFT + 1) \times h(l)/Sample\ rate \tag{7}$$

The filter bank can be computed as;

$$M_f(k) = \begin{cases} 0 & k < G(f-1) \\ \frac{k-G(f-1)}{G(f)-G(f-1)} & G(f-1) \le k \le G(f) \\ \frac{G(f+1)-k}{G(f+1)-G(f)} & G(f) \le k \le G(f+1) \\ 0 & k > G(f+1) \end{cases} \tag{8}$$

where $f = 1$ to F is Mel Filters number.

2.6 Filter bank energy

The filter bank is bonded by power spectrum and summed up to some coefficients. The filter bank energy can be computed as;

$$E(l) = \sum_{m=0}^{\frac{m}{2}} \log |A(m)|B(m)\left(k\frac{2\pi}{M}\right) \times WT_m \tag{9}$$

where WT_m is the multiple kernel weighted function that can be computed as;

$$WT_m = WT_{m1} + WT_{m2} \tag{10}$$

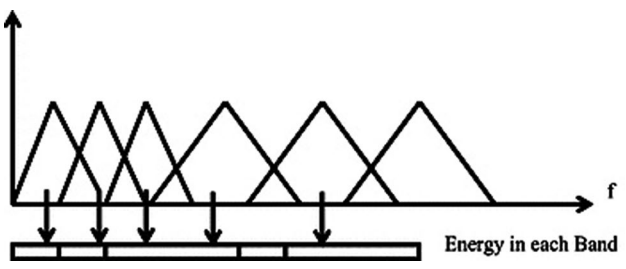

Fig. 2 Mel scale filter bank

2.7 Discrete cosine transform (DCT)

The DCT is performed for transforming the log Mel spectrum estimates to spatial domain.

$$E(l) = \overline{E}(k) \tag{11}$$

where

$$\overline{E}(k) = \begin{cases} E(l), & k = k_l \\ 0, & otherwise \end{cases} \tag{12}$$

The cepstral coefficient can be computed as;

$$WC_s(m) = \frac{1}{M'} \sum_{k=0}^{M'-1} \bar{E}(k)\, e^{jk(2\pi/M')m} \tag{13}$$

$WC_s(m)$ represents multiple kernel weighted Mel frequency cepstral coefficients.

2.8 Delta energy and spectrum

The energy patterns or features are summed within the acoustic features vector. The addition enhances audio recognition accuracy and dominates noise robustness as well as the echo.

2.9 Cepstral normalization

In the normalization procedure, the average of each of coefficients will be subtracted and divided with variance.

3 Classification using support vector machine

The classification stage in ASI systems is a feature matching procedure among the new speaker features and the database saved features. The SVM depends on the statistical learning theory (Boujelbene et al. 2010). It is based on finding the best interval among between feature levels to be precisely isolated as much as possible. Such features must be divided linearly using the hyper-plane which may be consider like linear classifier. The SVM transform input features into feature space with large dimension (Zergat and Amrouche 2014; Campbell et al. 2007).

3.1 Geometric margin

It is required to estimate the space from the two patterns to separator. The space is the margin among the two patterns

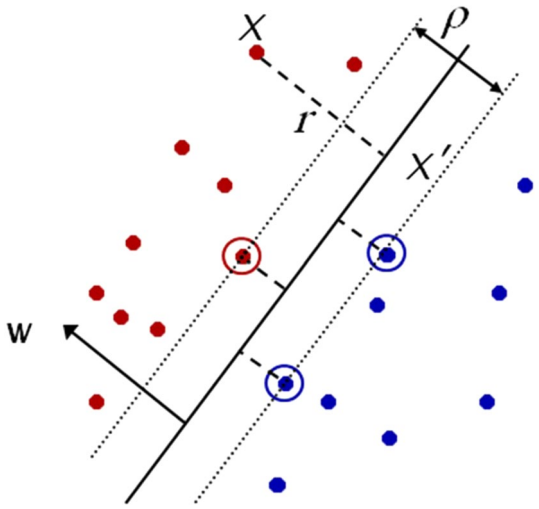

Fig. 3 Separating different patterns with a hyperplane

which is the minimum space among the pattern and hyperplane, defined with dashed line in Fig. 3.

For formulating such distance r, let x′–x defines the dotted line which is perpendicular to decision border and parallel to the hyperplane with the normal vector w. The unit vector of normal vector direction to the hyperplane may be estimated as:

$$\bar{v} = \frac{w}{|w|} \tag{14}$$

So that the distance r may be estimated as:

$$\bar{r} = r \star \bar{v} \tag{15}$$

Since,

$$r = X' - X \tag{16}$$

So,

$$X' = X - r\frac{w}{|w|} \tag{17}$$

The margin among the hyper-plane and the closest two patterns of the two data classes may be estimated as:

$$z(X_i) = y_i(w^T X_i + b) \tag{18}$$

where w is the decision hyperplane normal vector, X_i is the data point, and y_i is the data point class (+1 or −1).

The margin distance may be estimated as:

$$\rho = \frac{2}{\|w\|} \tag{19}$$

3.2 Separation technique of SVM

The main aim of SVM is to determine the optimal separately hyperplane. So, the optimal separately hyperplane may be considered as optimizing problem:

$$\begin{aligned} &\text{Maximize} : \rho \\ &\text{Subject to} : z(X_i) \geq 1 \end{aligned} \tag{20}$$

Using Lagrange multiplier scheme, Eq. (20) can be minimized and the objective function can be restated as:

$$L(w,b,\alpha) = \frac{1}{2}ww - \sum_{i=1}^{n} \alpha_i(y_i(wx_i + b) - 1) \tag{21}$$

where constant α_i is Lagrange multiplier. By differentiating α_i with respect to w and b:

$$\frac{\partial L(w,b,\alpha)}{\partial w} = w - \sum_{i=1}^{n} \alpha_i y_i x_i = 0 \tag{22}$$

$$\frac{\partial L(w,b,\alpha)}{\partial b} = \sum_{i=1}^{n} \alpha_i y_i = 0 \tag{23}$$

Substituting from Eqs. (22) and (23) into Eq. (21):

$$L(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j x_i x_j \tag{24}$$

The minimization of Eq. (24) can be considered as a convex quadratic programming problem with condition:

$$\sum_{i=1}^{n} y_i \alpha_i = 0 \text{ and } \alpha_i \geq 0 \tag{25}$$

The minimization of the Eq. (24) will be:

$$L(w,b,\alpha) = \sum_{i=1}^{n} \alpha_i - \frac{1}{2}\sum_{i,j=1}^{n} y_i y_j \alpha_i \alpha_j k(X_i,X_j) \tag{26}$$

The hyperplane may be estimated as:

$$z(X) = \text{sign}\left[\sum_{i}^{n} y_i \alpha_i k(X_i,X_j) + b\right] \tag{27}$$

4 The proposed MKMFCC–SVM ASI system

The full description of the proposed MKMFCC–SVM ASI system using MKMFCC feature extraction and SVM classification algorithm is addressed. Initially, the audio signals for multiple speakers are taken as input for the ASR system. Feature extraction is performed in which the feature vector sequences representing feature patterns about speech signal is extracted. The MFCC features are extracted, and multiple kernel weighted function is performed for generating the MKMFCC coefficients using Mel filter bank energy.

After feature extracting phase, speech classification phase is employed with the SVM.

4.1 Feature extraction phase

The feature extraction phase include speaker related properties for effective recognition. The KMFCCs are considered within the proposed ASI since it enhances and preserves information formant from spectral envelope. The MFCC spectral feature differs from other acoustic features in time frequency analysis and requency smoothing schemes.

4.2 SVM implementation for feature matching phase

The research paper utilizes sequential minimal optimization (SMO) (You et al. 2010). The SMO selection rather than other optimization schemes is due to reliability of SMO scheme on large datasets and the LIBSVM library utilized for SVM implementation using SMO can be linked to the Matlab platform. Much time is required for Kernel matrix calculation utilized in SVMs under normal situation, this time grows quickly when training samples number are exist, resulting in a larger Kernel matrix. To bypass such difficulty, SMO divides the problem into a series of smaller quadratic programming problems. The SMO procedure may be summarized as:

Step 1 Choose an arbitrary Lagrange multiplier α.
Step 2 Choose other Lagrange multiplier.
Step 3 Upgrade the other second Lagrange multiplier using Eq. (28):

$$\alpha_2^{new} = \alpha_2 + \frac{y_2(E_1 - E_2)}{k} \tag{28}$$

Step 4 Set the Lagrange multiplier, i.e. $\alpha_2^{\text{new, assigned}} \leftarrow \alpha_2^{\text{new}}$.
Step 5 If the Lagrange multiplier is not varied, go back to Step 1.
Step 6 Upgrade the earliest Lagrange multiplier.
Step 7 If all Lagrange multiplier satisfy step 5 conditions, end. Else, go to step 1.

5 Experimental tests

With existence of telephone and noise-analogous degradations, speaker recognition process may be not an easy process. The noise-analogous degradation tries to disguise the speech signal so the extracted features will not accurate and infeasible for recognition. The telephone-analogous degradation may be considered as a low-pass filter on the speech signal that may remove a lot of speaker features. In this section, different four speaker recognition tests are performed with different degradation types. The considered degradations will be AWGN, colored noise, telephone-analogous degradations with AWGN and telephone-analogous degradations with colored noise. The telephone-analogous degradations have been simulated using low-pass filter of low bandwidth applied on speech signals.

During ASI training stage, a database that includes 80 speakers is utilized. Every speaker iterates a given Arabic clause 15 times. As a result, 1200 speech models will be utilized for providing MKMFCCs using the proposed MKMFCC–SVM ASI, MFCCs and polynomial coefficients for MFCC–SVM ASI to constitute database features vector. During enrolling stage, every speaker is requested to repeat the clause and the audio signal is subjected to degradation. Comparable features like utilized during enrollment will be also evolved from such the degraded speech signals, and utilized in the classification stage. Five methods for feature extraction are employed in the paper.

In first scheme, features extraction of the MKMFCCs, and MFCCs is performed directly using only the speech signals. In second scheme, features extraction is performed using DCT of speech signals. In third scheme, features extraction is performed from the concatenation of both the original speech signal and DCT of speech signal in one features vector. In fourth scheme, features extraction is performed using DWT of speech signals. In fifth scheme, features extraction is performed from the concatenation of both the original speech signal and DWT of speech signal in one features vector. Comparisons are performed to inspect the performance of MKMFCC–SVM ASI with respect to MFCC–SVM ASI in terms of identification rate using the above mentioned five feature extraction schemes in four degradation situations, and test results are shown in Tables 1, 2, 3 and 4. Firstly, the results shown in Tables 1, 2, 3 and 4 ensured and proved the superiority of the proposed MKMFCC–SVM ASI compared with MFCC–SVM ASI using the five feature extraction schemes in all the four degradation cases. Also, it is clear from the results in Tables 1, 2, 3 and 4 for both the proposed MKMFCC–SVM ASI and MFCC–SVM ASI that the extracted features from the audio plus DWT audio signals and audio plus DCT audio signals have the highest recognition rate in all the four degradation cases. In AWGN case, the extracted features using speech plus DWT speech signals have the best recognition rates with different SNRs. For colored noise case shown in Table 2, the extracted features using speech plus DCT speech signals achieve the best recognition rates with different SNRs. In telephone-analogous degradations with AWGN and colored noise cases shown in Tables 3 and 4, respectively, the performance suffers since the low-pass filter eliminates a lot of speech features. The extracted features using speech plus DWT speech signals achieve the best recognition rates for the telephone like degradations and AWGN at all SNRs.

Table 1 Identification rate of MKMFCC–SVM ASI and MFCC–SVM ASI with different transforms in the presence of AWGN with different SNR

SBR (dB)	Audio features		Audio DCT features		Audio features + audio DCT features		Audio DWT features		Audio features + audio DWT features	
	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM
−20	50.5	47.25	54.5	52	56	53.25	59	56.5	60.25	58
−15	58.75	56.5	63	59.75	65	62.5	68.25	64.75	70.25	66.25
−10	66.5	63	67.5	65.75	69	67	70.5	68.75	72.75	69.5
−5	70	65.75	71.75	69.25	73.5	71.25	76	72.75	78.25	74.5
0	72.25	68	73.75	70.25	75	72.5	78	75	80.5	76.5
5	74	72.25	76.25	74.25	78.75	75.5	81.5	78.5	83	80
10	77.5	75	79.5	76.5	81.25	78	83.5	80.75	86.75	82.25
15	84.75	78.5	81	79.75	83.25	81.75	87.75	83.25	89.5	84.75
20	87	84.25	91.5	89.75	93.75	92.5	95.5	92.25	97.25	94

Table 2 Identification rate of MKMFCC–SVM ASI and MFCC–SVM ASI with different transforms in the presence of colored noise with different SNR

SBR (dB)	Audio features		Audio DCT features		Audio features + audio DCT features		Audio DWT features		Audio features + audio DWT features	
	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM
−20	49.25	46	58	55	59	56.75	53.25	51	54.75	52
−15	57.5	55	67.25	63.5	69.25	65	62	58.5	64	61.25
−10	65.25	61.75	69.5	67.75	71.5	69	66.5	64.5	68	66
−5	69	64.5	75	71.5	77	73.25	70.5	68	72.25	70
0	71	67	77.25	74	79.5	75.25	72.5	69.25	74	71.25
5	73.25	71.5	80.25	77.25	82	79	75	73	77.5	74.25
10	76	73.75	82.5	79.5	85	81	78.5	75.25	80	77
15	83.25	77.5	86.75	82	88.25	83.5	80	78.75	82.25	80.25
20	86	83	94.25	92.25	96.25	92.75	90.25	88.25	92.5	91

Table 3 Identification rate of MKMFCC–SVM ASI and MFCC–SVM ASI with different transforms in the presence of telephone like degradation and AWGN with different SNR

SBR (dB)	Audio features		DCT audio features		Audio features + DCT audio features		DWT audio features		Audio features + DWT audio features	
	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM
−20	48	45.5	52.25	50.5	53.5	51.25	57.5	54.25	58.75	56.75
−15	57	55	61.25	58.5	61.5	60.25	66.25	62.5	67.75	64.5
−10	64	61.5	66	64.5	67.25	65.5	69.75	66.5	71.5	67.75
−5	67.5	64	69.5	67.25	71.25	68.5	74.25	70.5	76.5	72.25
0	70.5	66.5	72	70.25	73.75	71.5	76.25	73.75	79.25	74.75
5	72	70.5	75	73.5	76.5	74.75	79.75	76.25	80.25	77.5
10	75	73.5	77.25	75	78.75	76.25	82.25	78.5	84.5	79.75
15	83.5	76.25	79.5	77.25	80.75	78.5	85.5	80	87.5	82.25
20	85.25	82.5	89.75	88.25	91	89.5	93.25	90.5	94.5	92.25

Table 4 Identification rate of MKMFCC–SVM ASI and MFCC–SVM ASI with different transforms in the presence of telephone like degradation and colored noise with different SNR

SBR (dB)	Audio features		DCT audio features		Audio features + DCT audio features		DWT audio features		Audio features + DWT audio features	
	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM	MKMFCC–SVM	MFCC–SVM
−20	46.5	44.25	56.25	53.5	57.5	54.25	51.5	49.25	52.25	50.25
−15	55.25	54.25	63.5	60.75	66.5	62.75	60.25	58.75	60.5	59.5
−10	61.25	60	67.5	65	70.25	66.5	64.5	62.75	66.75	64.25
−5	65.75	62.5	72.75	69.75	74.75	70.5	68.25	66.25	70.5	67.75
0	69.75	64.25	75	72.25	77.5	73.5	71	69.5	72.5	70
5	71.5	69.5	78	75	79.5	75.75	73.75	71.75	74.75	72.5
10	73	71.25	81.5	77.25	83	78.5	76.5	73.5	77.75	74.75
15	82.5	75.5	84	79.25	86	81	78.25	75.25	80	76.75
20	84.5	80.25	92.5	88.75	93.25	90.5	88.25	86.5	90.75	88.25

But the extracted features from the audio plus DCT audio signals achieve the best recognition rates for the telephone like degradations and colored noise at all SNRs.

6 Conclusion

The paper introduced an efficient robust noise ASI method using MKMFCC and SVM. A comparative study is held between the proposed MKMFCC–SVM ASI and MFCC–SVM ASI in terms of identification rate measure using five methods for extracting features in presence of five degrading cases. Experimental tests prove the effectiveness of the proposed MKMFCC–SVM ASI for extracting features from telephone and noisy-like degraded audio signals.

References

Boujelbene, S. Z., Mezghani, D. B. A., & Ellouze, N. (2010). Improving SVM by modifying kernel functions for speaker identification task. *International Journal of Digital Content Technology and its Applications, 4(6*, 100–105.

Campbell, W. M., Campbell, J. P., Gleason, T. P., Reynolds, D. A., & Shen, W. (2007). Speaker verification using support vector machines and high-level features. *IEEE Transactions on Audio, Speech and Language Processing, 15*(7), 2085–2094.

Dharanipragada, S., Yapanel, U. H., & Rao, B. D. (2007) Robust feature extraction for continuous speech recognition using the MVDR spectrum estimation method. *IEEE Transactions on Audio, Speech, and Language Processing, 15*(1), 224–234.

Ding, I.-J., & Yen, C.-T. (2015) Enhancing GMM speaker identification by incorporating SVM speaker verification for intelligent web-based speech applications. *Multimedia Tools and Applications, 74*, 5131–5140.

Furui, S. (1981) Cepstral Analysis Technique for Automatic Speaker Verification. *IEEE Transactions on Acoustics, Speech, and Signal Processing, 20*(2), 254–272.

Galushkin, A. I. (2007). *Neural networks theory*. Berlin: Springer.

Gandhiraj, R., Sathidevi, P. S. (2007). Auditory-based wavelet packet filter bank for speech recognition using neural network. In *Proceedings of the 15th International Conference on Advanced Computing and Communications*, pp. 666–671.

Hayati, M., shirvany, Y. (2007). Artificial neural network approach for short term load forecasting for Illam region. *Proceeding of World Academy of Science, Engineering and Technology*, 22. ISSN 1307–6884.

Hossain, M., Ahmed, B., Asrafi, M. (2007). A real time speaker identification using artificial neural network. In *10th International Conference on Computer and Information Technology*, pp. 1–5.

Huang, C., Song, B., & Zhao, L. (2016). Emotional speech feature normalization and recognition based on speaker-sensitive feature clustering. *International Journal of Speech Technology, 19*, 805–816.

Li, Z., & Gao, Y. (2016). Acoustic feature extraction method for robust speaker identification. *Multimedia Tools and Applications, 75*, 7391–7406.

Mellahi, T., & Hamdi, R. (2015). LPC-based formant enhancement method in Kalman filtering for speech enhancement. *International Journal of Electronics and Communications, 69*(2), 545–554.

Naeeni, B. H., Amindavar, H., & Bakhshi, H. (2010). Blind per tone equalization of multilevel signals using support vector machines for OFDM in wireless communication. *International Journal of Electronics and Communications, 64*(2), 186–190.

Polur, P. D., & Miller, G. E. (2005). Experiments with fast Fourier transform, linear predictive and cepstral coefficients in dysarthric speech recognition algorithms using hidden Markov model. *IEEE Transactions on Neural Systems and Rehabilitation Engineering, 13*(4), 558–561.

Qian, F., Hu, G., & Yao, X. (2008). Semi-supervised internet network traffic classification using a Gaussian mixture model. *International Journal of Electronics and Communications, 62*(7), 557–564.

Ramaiah, V. S., & Rao, R. R. (2016). Speaker diarization system using MKMFCC parameterization and WLI-fuzzy clustering. *International Journal of Speech Technology, 19*, 945–963.

Selva Nidhyananthan, S., Shantha Selva Kumari, R., & Senthur Selvi, T. (2016). Noise robust speaker identification using RASTA-MFCC Feature with quadrilateral filter bank structure. *Wireless Personal Communications, 91*, 1321–1333.

Shuling, L., & Wang C. (2009). Nonspecific speech recognition method based on composite LVQ1 and LVQ2 network. In *Chinese Control and Decision Conference (CCDC)*, pp. 2304–2388.

Xu, L., & Yang, Z. (2016). Speaker identification based on state space model. *International Journal of Speech Technology, 19*, 404–414.

You, C. H., Lee, K. A., & Li, H. (2010). GMM-SVM kernel with a Bhattacharyya-based distance for speaker recognition. *IEEE Transactions on Audio, Speech, and Language Processing, 18*(6), 1300–1312.

Zergat, K. Y., & Amrouche, A. (2014). New scheme based on GMM-PCA-SVM modeling for automatic speaker recognition. *International Journal of Speech Technology, 17*, 373–381.