Speaker Verification with Multi-Run ICA Based Speech Enhancement

Ahmed Kamil Hasan AL-ALI, David Dean, Bouchra Senadji, Mahsa Baktashmotlagh, and Vinod Chandran
School of Electrical Engineering and Computer Science, Queensland University of Technology, Brisbane, Australia
Email: ahmedkamilhasan.alali@hdr.qut.edu.au, ddean@ieee.org,{b.senadji, m.baktashmotlagh}@qut.edu.au

Abstract*—Forensic speaker verification systems show severe performance degradation in the presence of noise when the signal to noise ratio (SNR) is low. A possible solution to this problem is the use of multi-run independent component analysis (ICA) to reduce the effect of noise from the noisy speech signals. Previous works have used multi-run ICA in biosignal application; however, the effectiveness of multi-run ICA on noisy speaker verification has not been investigated yet. In this paper, we use multi-run ICA to enhance the noisy speech signals by choosing the highest signal to interference ratio (SIR) of the mixing matrix from different mixing matrices generated by iterating the fast ICA algorithm for several times. We use a combination of feature-warped mel frequency cepstral coefficients (MFCCs) and feature-warped MFCC extracted from the discrete wavelet transform (DWT) of the enhanced speech signals as the feature extraction. A state-of-the-art identity vector (i-vector) probabilistic linear discriminant analysis (PLDA) was used as a classifier in this paper. Experimental results demonstrate that the proposed method with multi-run ICA achieves high improvements in equal error rate (EER) of 66.68%, 69.24% and 70.78% over the baseline noisy speaker verification system, when the test speech signals are corrupted with CAR, STREET, and HOME noises respectively at -10 dB SNR.

I. Introduction

A current issue in the field of speaker verification systems is to migrate automatic speaker verification developed in the lab to real forensic situations. The forensic audio recordings from the suspect are often recorded using hidden microphones in public places [1]. Such forensic audio recordings are often corrupted with car and street noises [1], [2]. The distortion of speech by environmental noise degrades significantly the high performance of speaker verification in the lab [3]. Therefore, speech enhancement in real forensic applications plays an important role in improving forensic speaker verification performance under noisy conditions.

Speech enhancement algorithms can be classified into single and multiple-channel algorithms based on the number of the microphones used to collect noisy speech signals. Various algorithms for single-channel speech enhancement have been proposed in previous studies [4], [5], [6], but single-channel speech enhancement algorithms achieve less improvement in speech quality compared to multiple-channel speech enhancement algorithms [7].

Independent component analysis was used in the multi-channel speech enhancement algorithm to separate the noise from the noisy speech signals [8], [9], [10]. It transforms the mixed signals into components that are statistically independent. The principle of estimating the independent components is based on maximizing the non-Gaussian distribution of one of the source signals [11]. The difference between the Gaussian distribution and the distribution of the independent component can be estimated using various non-Gaussian measurements such as Kurtosis, negentropy, and approximation of the negentropy [11].

There is a randomness in the separation of the mixed signals in the ICA algorithm due to the randomness associated with choosing the unmixing matrix [12]. The randomness of separation of the source signals in the ICA algorithm may decrease the performance of recognition systems. The multi-run ICA algorithm [13] was proposed to deal with this problem and it achieved better performance for biosignal applications under clean conditions. This method is based on iterating the fast ICA algorithm [11] several times. Different mixing matrices were obtained each time. The source signals can be estimated by choosing the best of the mixing matrix. Selecting the best mixing matrix can be performed by computing the optimum SIR of the mixing matrices [13].

A number of techniques have been used to model acoustic speech for speaker verification such as Gaussian mixture model based on universal background model (GMM-UBM) [14], support vector machine (SVM) [15] and joint factor analysis (JFA) [16]. Recently, the identity vector (i-vector) [17] feature has become more popular technique for speaker verification. The i-vector represents the GMM super-vector by using a single total variability matrix. This single total variability was motivated by the discovery that speaker information in channel space of JFA could be used to recognize between speakers more efficiently [18]. The i-vector can be used in Gaussian probabilistic linear discriminant analysis (GPLDA) and heavy-tailed PLDA (HTPLDA) [19]. It was found in [19] that HTPLDA achieved better speaker verification performance compared with GPLDA. Recently, the length-normalized GPLDA technique [20] has proposed. This technique is based on transforming the distribution of i-vector from the heavy-tailed to Gaussian distribution. The results in [20] have indicated that the length-normalized GPLDA achieved similar performance with less computational complexity compared with HTPLDA. Thus, the length-normalized GPLDA will be used in this paper.

In this work, we present approach for noisy forensic speaker verification systems. This approach uses the multi-run ICA algorithm to reduce the effect of environmental noise from the noisy speech signals. Combination features of DWT-based feature-warped MFCC and feature-warped MFCCs of the enhanced speech signals are used as the feature extraction. These features can be used to train the state-of-the-art length-normalized GPLDA speaker verification systems.

The effectiveness of the multi-run ICA algorithm on reducing the effect of environmental noise and improving the performance of noisy speaker verification has not been investigated yet in the literature. This algorithm can be used for improving the performance of noisy forensic speaker verification because it is based on choosing the best of the unmixing matrix to separate the noise from the noisy speech signals; this is the original contribution of this research.

The remainder of the paper is structured as follows: Section II describes the speech corpus. The model of ICA is presented in Section III. Section IV describes the methodology of speaker verification in the presence of noise. The results and discussion are given in Section V. Finally, Section VI concludes the paper.

II. Speech corpus

The speaker verification system was evaluated using the Australian forensic voice comparison database [21]. This database consists of 552 speakers and recorded in three speaking styles for each speaker [21]. The speaking styles used in this database include: informal telephone conversations, information exchange over the telephone and pseudo-police styles [21]. Informal telephone conversations and information exchange over the telephone are recorded between two speakers using a telephone channel. The pseudo-police interview style is recorded using a microphone channel. The clean speech signals are sampled at 44.1 kHz and 16 bit/sample resolution [22].

The enrolment of the speech signals was obtained by using full duration utterances from 200 speakers from the pseudo-police interview style. The test speech signals were obtained using 10 second durations from 200 speakers in informal telephone conversation styles. Voice activity detection (VAD) based on the statistical model developed by Sohn *et al.* [23] was used to remove the silence frames from the enrolment and test speech signals.

III. Model of independent component analysis

In most real forensic scenarios, the criminal may use a mobile to commit criminal offenses. The test speech signals from forensic audio recordings are often corrupted by various types of environmental noise in open areas such as street. Thus, the effect of reverberation on noisy test speech signals will not be investigated in this paper. We used instantaneous ICA to estimate the enhanced speech from the noisy speech signals. Let the speech and noise signals emitted from n sources be represented as $s(t) = \{s_1(t), s_2(t), ..., s_n(t)\}$. The mixed speech signals can be recorded instantaneously in public places for forensic applications by using m microphones and be expressed as $x(t) = \{x_1(t), x_2(t), \cdots, x_m(t)\}$ [1].

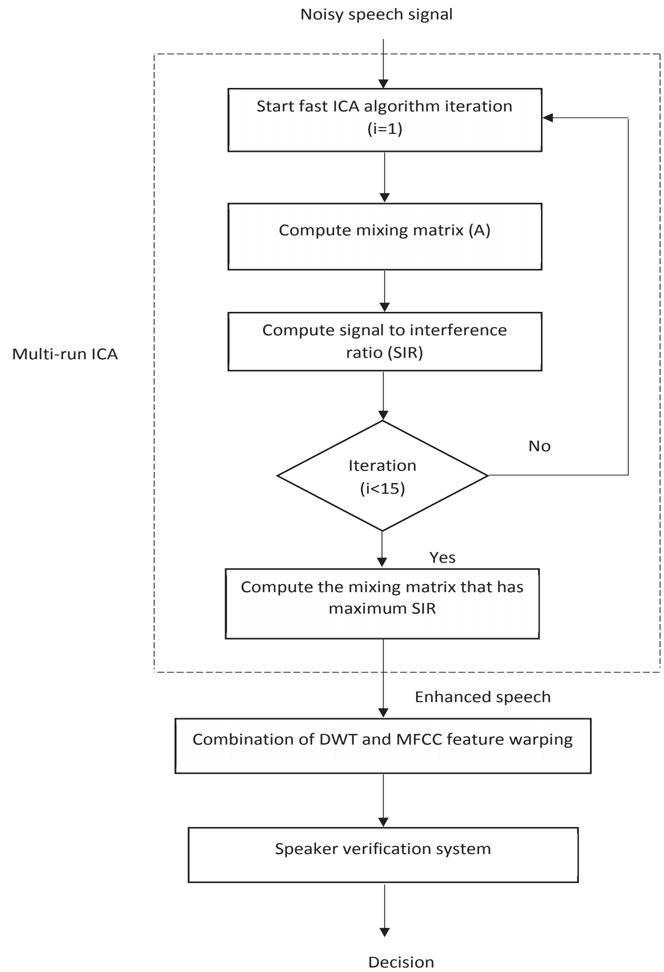

Fig. 1. Speaker verification based on multi-run ICA.

Instantaneous ICA assumes the mixing process as linear, so it can be represented as [11]

$$x = As \tag{1}$$

where A is $m \times n$ mixing matrix

The goal of ICA is to estimate the source signals from the mixed signals when both the input sources and the parameters of the mixing matrix are unknown. The estimates of source signals ($\hat{s}$) can be represented by the following equation [11]:

$$\hat{s} = Wx \tag{2}$$

where W is the unmixing matrix. The unmixing matrix equals to the inverse of the mixing matrix A and the ICA assumes the number of the mixed signals equal to the number of the source signals [24]. In this paper, we use two source signals (speech and noise) and two microphones to record the noisy speech signals ($m = n = 2$). Thus, the mixing and unmixing matrices are square and they have a size of 2×2.

IV. Methodology

Experiments were conducted to evaluate the performance of the forensic speaker verification system in the presence of environmental noise. The system of speaker verification based on multi-run ICA consists of the following steps, which are shown in Figure 1 and described in the next sections.

A. Noisy speech signals

Each of the test speech signals was mixed with one random session of the environmental noises (CAR, STREET, and HOME noise) from the QUT-NOISE database [25], resulting in a two-channel noisy speech signal. These noises were downsampled from 48 kHz to 44.1 kHz before mixing with the clean speech signal. The down sampling is necessary to match the sampling frequencies of the clean speech and noise signals.

Figure 2 shows configuration of sources ($s(n)$ and $e(n)$) and microphones (x_1 and x_2) in instantaneous ICA. The noisy speech signals can be recorded by the microphones, x, as

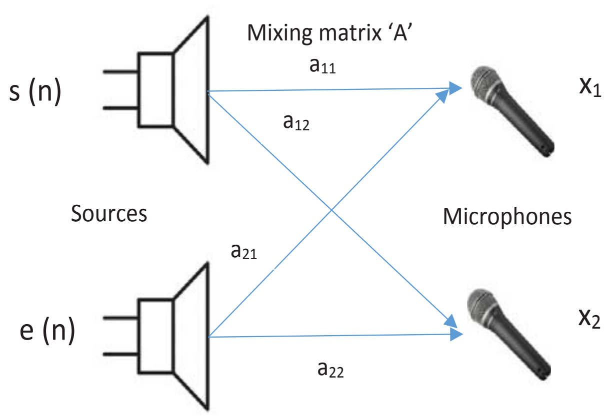

Fig. 2. Configuration of sources and microphones in instantaneous ICA.

$$x = As(n) \tag{3}$$

$$\begin{bmatrix} x_1 \\ x_2 \end{bmatrix} = \begin{bmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{bmatrix} \begin{bmatrix} s(n) \\ e(n) \end{bmatrix} \tag{4}$$

where $s(n)$ is the clean speech signal and $e(n)$ is the environmental noise. The mixing matrix used in this paper is

$$A = \begin{bmatrix} 1.0 & 1.0 \\ 1.0 & 0.6 \end{bmatrix} \tag{5}$$

B. Fast ICA

The procedure for a fast ICA algorithm for one unit can be illustrated by the following steps [11]:

1) Remove the mean value from the noisy signal and center its distribution.
2) Whiten the noisy speech signal (x) to get (x_w) by using eigenvalue decomposition of the covariance of the noisy speech signal.
$$x_w = VD^{-1/2}V^T x \tag{6}$$
where V is the eigenvector matrix of the covariance of the noisy speech signal, and $D^{-1/2}$ is the inverse square root diagonal of the eigenvalue matrix.
3) Choose an initial vector of unmixing matrix.
4) Estimate a row vector of unmixing matrix
$$w^+ = E\{x_w\, g(w^T x_w)\} - E\{g'(w^T x_w)\}w \tag{7}$$
where w^+ is the new value of the row vector of the unmixing matrix, E is the sample mean, g and g' are the first and the second derivatives of the contrast function respectively. The Gaussian contrast function was used in this paper and it can be defined as [26]
$$G(u) = -exp(-au^2/2) \tag{8}$$
where a is a constant and it equals to one.
5) Normalize the row vector of w^+
$$w^* = \frac{w^+}{\|w^+\|} \tag{9}$$
where w^* is the normalization of the new row vector of the unmixing matrix.
6) Insert $w = w^*$ in step 4 and repeat the procedure until there is convergence.

The criterion of convergence is that the direction of previous and new values of w must be in the same direction, i.e. the dot product of these w points is almost equal to one [11]. It is necessary to run one unit fast ICA algorithm for n times to estimate all source signals. To prevent different row vectors of w from converging to the same maxima, a deflation decorrelation of the output $w_1^T x, w_2^T x, \cdots, w_n^T x$ must be performed after every iteration [11].

C. Multi-run ICA

Independent component analysis is an iterative blind source separation approach [12]. The original source signals can be estimated from the mixing matrix A each time. The quality of the estimation of the source signals is based on the unmixing matrix W. Since the unmixing matrix is estimated from the random matrix, there is a randomness in the quality of the separation of the estimated source signals [12].

The multi-run ICA algorithm [13] deals with this problem of randomness. This algorithm is based on computing the mixing matrices several times and choosing the unmixing matrix that has the maximum quality of separation. In this paper, we iterate the fast ICA algorithm [11] for 15 times. Different mixing matrices $A_1, A_2, .., A_{15}$ are obtained from the fast ICA iteration. To estimate the enhanced speech from the noisy speech signals, we require just one mixing matrix. Thus, the best of the mixing matrix has to be chosen among different mixing matrices in order to obtain better speaker verification performance. There are various methods to compute the quality of the separation of the unmixing matrix. For audio and biomedical applications, the SIR was found to be the more popular method to compute the quality of the separation [27] and is used in this paper. The SIR is the ratio of the power of the wanted signal to the power of the unwanted signal and it can be represented as [28]

$$SIR = \frac{||S_{target}||^2}{||e_{interf}||^2} \tag{10}$$

where S_{target} is the modified version of the clean speech signal $s(n)$ by an allowed distortion and e_{interf} is the interference error. The enhanced speech signal can be estimated by choosing the unmixing matrix that has the maximum SIR.

D. Combination of DWT and MFCC feature warping approach

The approach of extracting the features is based on the multi-resolution property of the DWT. The MFCCs were computed over Hamming windowed frames of the enhanced speech signals with 30 ms size and 10 ms shift. The MFCCs were obtained using a mel filterbank of 32 channels, followed by a transformation to the cepstral domain, keeping 13 coefficients. The first and second derivatives of the cepstral coefficients were appended to the MFCCs. Feature warping with a 301 frame window is applied to the features extracted from the MFCCs. The frame of the enhanced speech signal is decomposed into two frequency sub-bands: low-frequency sub-band (approximation coefficients) and high frequency sub-band (detail coefficients). The approximation and detail coefficients are combined into a single vector. The decomposition process can be repeated by applying the DWT to the low-frequency sub-band. To capture the essential characteristics of the vocal tract, the feature-warped MFCCs are applied to the concatenated vector from the DWT. Finally, the combination of DWT and feature-warped MFCCs approach can be performed by combining the feature-warped MFCCs from the full band of the enhanced speech signal with MFCC feature warping extracted from the DWT in a single feature vector [29]. Figure 3 shows combination of DWT and MFCC feature warping technique. The symbol (FW) in Figure 2 represents the acronym of feature warping.

E. Speaker verification system

The speaker verification system was conducted with a state-of-the-art Gaussian PLDA (GPLDA) i-vector framework. It used a combination of DWT and feature-warped MFCCs as a front-end. The universal background model (UBM) with 256 components was used in the experiments. The UBM was trained on telephone and microphone speech using 348 speakers from the Australian forensic voice comparison database [21]. The UBM was used to calculate the Baum-Welch statistics for calculation of the total variability subspace of dimension 400. The total variability was used to compute the i-vectors. The i-vectors were projected into a linear discriminant analysis (LDA) space to reduce the dimension of i-vectors to 200. The i-vectors length normalisation was used before GPLDA modeling using centering and whitening of the i-vectors [20]. The performance of the i-vector PLDA speaker verification system was evaluated using the Microsoft research (MSR) identity toolbox [30].

V. Results and discussion

This section describes the performance of the speaker verification systems based on multi-run ICA under noisy conditions. The EER was used to evaluate the performance of speaker verification systems.

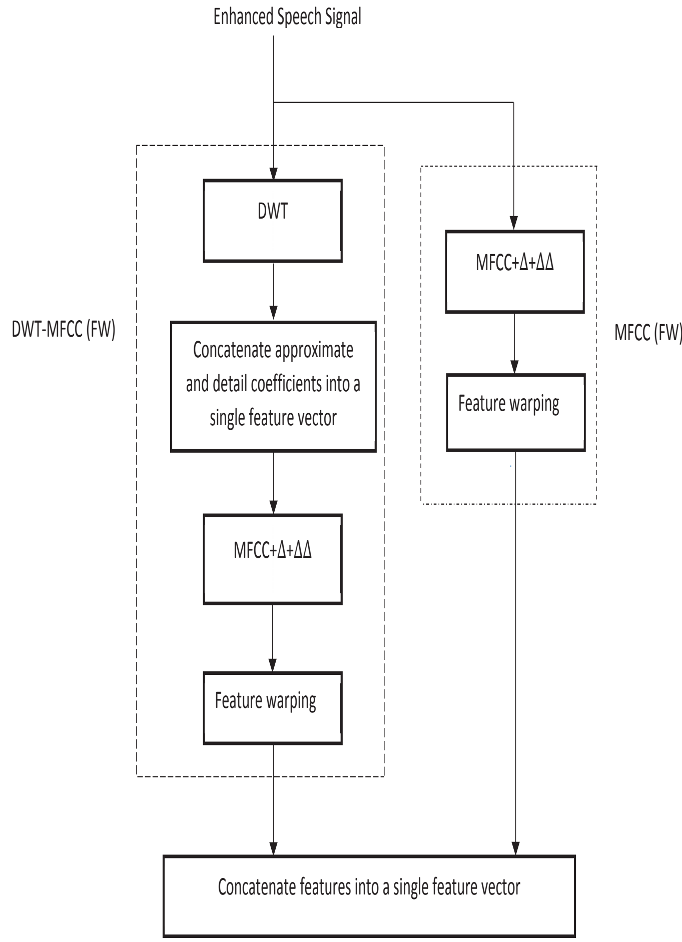

Fig. 3. Combination of DWT and MFCC feature warping.

A. Effect of level decomposition

This experiment evaluated the effect of level decomposition used in the performance of noisy speaker verification without using speech enhancement (baseline). The full duration of enrolment speech signals was kept in clean conditions, while 10 sec of the test speech signals were corrupted with a random session of STREET, CAR, and HOME noises at SNRs ranging from -10 dB to 10 dB. The enrolment and noisy test speech signals were decomposed into 2, 3, and 4 levels using Daubechies 8 DWT.

Figure 4 shows the effect of the decomposition levels on the performance of speaker verification without using speech enhancement in the presence of various types of environmental noise. It was found that increasing the number of levels to more than three over the majority of SNR values degraded the speaker verification performance in the presence of various types of environmental noise. In this case, the number of samples in the lowest frequency subbands was so low that the essential characteristics of the speech signals could not be estimated accurately by the classifier [31]. Thus, level 3 is used in the feature extraction based on combination of DWT

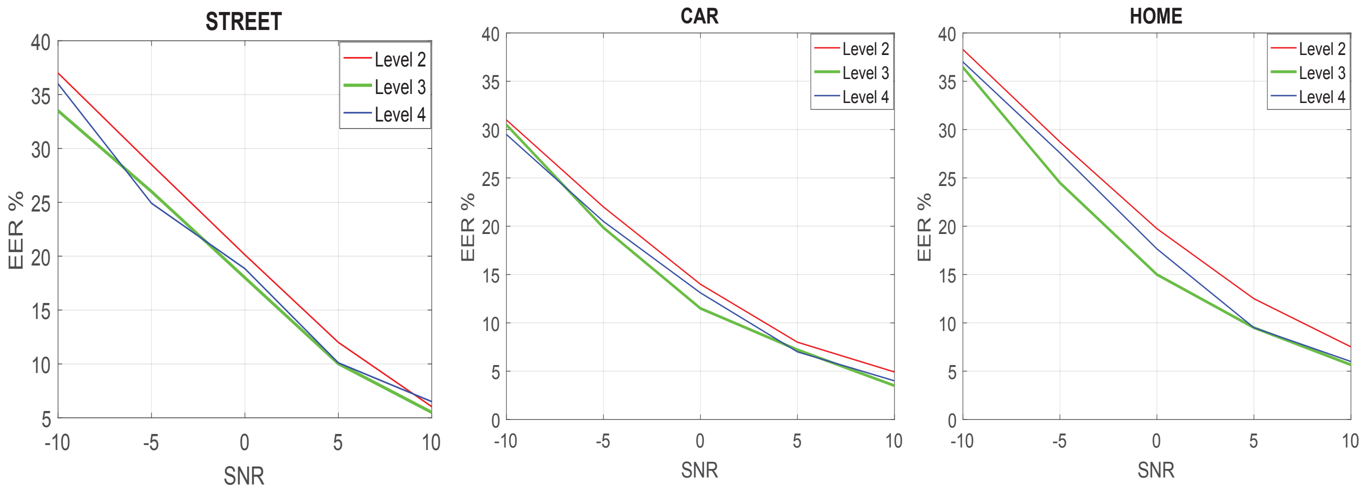

Fig. 4. Effect of the decomposition levels on the performance of speaker verification without using speech enhancement in the presence of environmental noise.

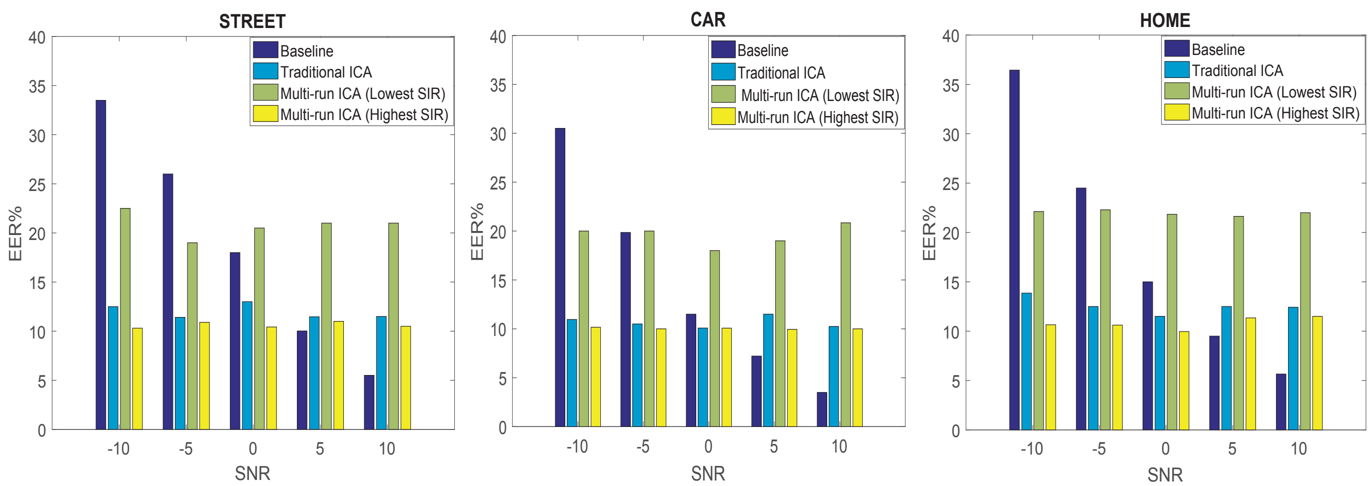

Fig. 5. Experimental results for PLDA speaker verification in the presence of STREET, CAR and HOME noises. Lower EER indicates better performance

and MFCC feature warping in the presence of noise in the next section.

B. Comparison of noisy speaker verification performance

This section describes the performance of speaker verification based on the multi-run ICA (highest SIR) and it compares with baseline speaker verification, traditional ICA and multi-run ICA (lowest SIR) in the presence of various types of environmental noises at SNRs ranging from -10 dB to 10 dB as shown in Figures 5. The SNRs on the x-axis of Figure 5 were computed from the first microphone (x_1).

It is clear from Figure 5 that the performance of speaker verification based on the multi-run ICA (highest SIR) algorithm achieved high improvements in EER over baseline noisy speaker verification at low SNR. The improvement in EER of speaker verification based multi-run ICA (highest SIR) decreased when SNR increased and there is a degradation in the speaker verification performance compared with the baseline noisy speaker verification in the presence of CAR, STREET and HOME noises at SNRs from 5 dB to 10 dB. The reduction in EER for speaker verification with multi-run ICA (highest SIR) was 66.68%, 69.24% and 70.78% over the baseline noisy speaker verification when the test speech signals were mixed with CAR, STREET and HOME noises respectively at -10 dB SNR. The multi-run ICA (highest SIR) algorithm achieved a significant reduction in EER at -10 dB SNR because this algorithm is based on choosing the best unmixing matrix that has the maximum SIR. The best matrix gives a clear separation of noise from the noisy speech signals.

The multi-run ICA with the lowest SIR degraded the noisy forensic speaker verification performance compared with traditional ICA. The estimation of the speech signals from the worst mixing matrix (lowest SIR) may occur at any instance in the traditional ICA algorithm due to the randomness associated with choosing the unmixing matrix. Thus, the traditional ICA algorithm may fail to separate the speech from the noisy speech signals in this case and this leads to decreased noisy forensic speaker verification performance.

The multi-run ICA (highest SIR) algorithm improved the forensic speaker verification performance over traditional ICA when the test speech signals were corrupted with different sessions of CAR, STREET and HOME noises at SNRs ranging from -10 dB to 10 dB. The reduction in EER for multi-run ICA over traditional ICA , EER_{red}, can be computed as

$$EER_{red} = \frac{EER_{ICA} - EER_{multi(ICA)}}{EER_{ICA}} \tag{11}$$

where EER_{ICA} is the equal error rate for traditional ICA and $EER_{multi(ICA)}$ is the equal error rate for multi-run ICA (highest SIR) approach. The average reduction in EER can be computed by calculating the mean in EER_{red} for various types of environmental noise at each noise level, as shown in Figure 6. The average reduction in EER for the multi-run ICA (highest SIR) algorithm ranges from 16.08% to 11.05% over conventional ICA in the presence of various types of environmental noise at SNRs ranging from -10 dB to 0 dB SNR.

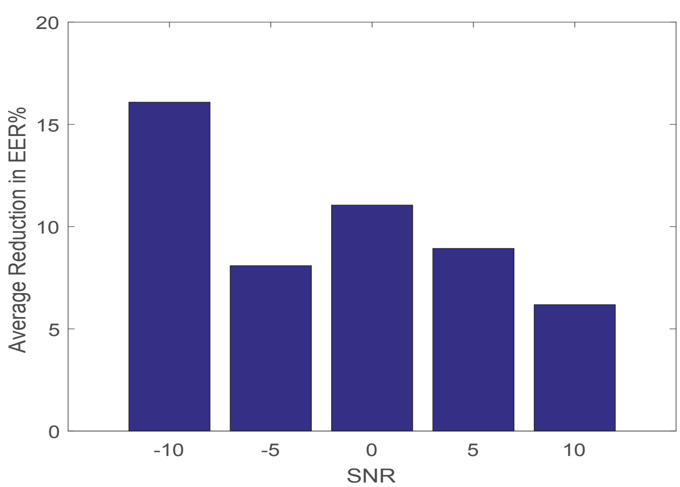

Fig. 6. Average reduction in EER for multi-run ICA algorithm over traditional ICA in the presence of various types of environmental noise at SNRs ranging from -10 dB to 10 dB. Higher average reduction in EER indicates better performance.

VI. Conclusion

In this paper, we investigated the effectiveness of the multi-run ICA algorithm to improve the performance of forensic speaker verification in the presence of environmental noise. This method is based on using the multi-run ICA algorithm to separate the noise from the noisy speech signals. Fusion of the DWT and feature-warped MFCCs was used to extract the features from the enhanced speech signals. Experimental results demonstrate that the method achieved great improvements in EER over the baseline noisy speaker verification when the test speech signals were mixed with various types of environmental noise at -10 dB SNR. These results indicate that the multi-run ICA algorithm can be very useful to improve the performance of forensic speaker verification in the presence of high levels of noise. Further work is required to investigate the effect of channel delay duration on the performance of noisy forensic speaker verification.

References

[1] A. K. H. Al-ALi, D. Dean, B. Senadji, V. Chandran, "Comparison of speech enhancement algorithms for forensic applications," in *16th Australian Int. Conf. Speech Sci. Technology*, December, 2016.

[2] M. I. Mandasari, M. McLaren and D. A. van Leeuwen, "The effect of noise on modern automatic speaker recognition systems," in *IEEE Int. Conf. Acoust., Speech Signal Process.*, 2012, pp. 4249-4252.

[3] D. Ribas, E. Vincent, and J. R. Calvo, "Full multicondition training for robust i-vector based speaker recognition," in *Interspeech*, Dresden, Germany, 2015.

[4] M. Berouti , R. Schwartz and J. Makhoul, "Enhancement of speech corrupted by acoustic noise", *IEEE Int. Conf. Acoust., Speech, Signal Process.*, vol. 4, 1979, pp. 208-211.

[5] D. L. Donoho and I. M. Johnston, "Ideal spatial adaptation by wavelet shrinkage", *Biometrika J.*, vol. 81, pp. 425-455,1994.

[6] K. Paliwal and A. Basu, "A speech enhancement method based on Kalman filtering," in *IEEE Int. Conf. Acoust., Speech, Signal Process.*, 1987, pp. 177-180.

[7] J. Rosca, R. Balan, and C. Beaugeant, "Multi-channel psychoacoustically motivated speech enhancement," in *Int. Conf. Multimedia Expo.*, 2003, pp. III- 217- III- 220.

[8] H. Y. Li, Q. H. Zhao, G. L. Ren and B. J. Xiao "Speech enhancement algorithm based on independent component analysis," in *5th Int. Conf. Natural Computation*, 2009, pp. 598-602.

[9] H. Li, H. Wang and B. Xiao, "Blind separation of noisy mixed speech signals based on wavelet transform and independent component analysis," in *8th Int. Conf. Signal Process.*, 2006.

[10] N. Shanmugapriya and E. Chandra, "Evaluation of sound classification using modified classifier and speech enhancement using ICA algorithm for hearing aid application," *ICTACT J. Commun. Technology*, vol. 7, pp. 1279-1288, March, 2016.

[11] A. Hyvärinen and E. Oja, "Independent component analysis: algorithms and applications", *Neural Netw.*, vol. 13, no. 4, pp. 411-430, 2000.

[12] G. R. Naik and D. K. Kumar, "Identification of hand and finger movements using multi run ICA of surface electromyogram," *J. Medical Syst.*, vol. 36, pp. 841-851, 2012.

[13] G. R. Naik, D. K. Kumar and M. Palaniswami, "Multi run ICA and surface EMG based signal processing system for recognising hand gestures," in *8th IEEE Int. Conf. Comput. Inform. Technology*, 2008, pp. 700-705.

[14] D. Reynolds, T. Quatieri, and R. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Process.*, vol. 10, no. 1-3, pp. 19-41, 2000.

[15] W. Campbell, J. Campbell, D. Reynolds, E. Singer, and P. TorresCarrasquillo, "Support vector machines for speaker and language recognition," *Comput. Speech Language*, vol. 20, pp. 210-229, 2006.

[16] P. Kenny, "Joint factor analysis of speaker and session variability: Theory and algorithms," tech. rep., CRIM, 2005.

[17] N. Dehak, P. J. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio, Speech, Language Process.*, vol. 19, pp. 788-798, 2011.

[18] N. Dehak, R. Dehak, P. Kenny, N. Brmmer, P. Ouellet, and P. Dumouchel, "Support vector machines versus fast scoring in the low-dimensional total variability space for speaker verification," *in Proc. Interspeech*, vol. 9. 2009, pp. 1559-1562.

[19] P. Kenny, "Bayesian speaker verification with heavy-tailed priors," in *Proc. Odyssey Speaker Lang. Recognit. Workshop*, 2010, pp. 1-10.

[20] D. Garcia-Romero and C. Espy-Wilson,"Analysis of i-vector length normalization in speaker recognition systems," in *Int. Conf. Speech Commun. Technology*, pp. 249-252, 2011.

[21] G. S. Morrison, C. Zhang, E. Enzinger, F. Ochoa, D. Bleach, M. Johnson, B. Folkes, S. De Souza, N. Cummins, and D. Chow, (2015). *Forensic database of voice recordings of 500+ Australian English speakers*. Available: http//databases.forensic-voice-comparison.net/.

[22] G. S. Morrison, P. Rose and C. Zhang, "Protocol for the collection of databases of recordings for forensic-voice-comparison research and practice," *Australian J. Forensic Sci.*, vol. 44, pp. 155-167, 2012.

[23] J. Sohn, N. S. Kim and W. Sung, ”A statistical model-based voice activity detection,” *IEEE Signal Pocess. Lett.*, vol. 6, no.1, pp. 1-3, Jan. 1999.

[24] P. Comon, ”Independent component analysis, a new concept?,” *Signal process.*, vol. 36, pp. 287-314, 1994.

[25] D. B. Dean, S. Sridharan, R. J. Vogt, and M. W. Mason, ”The QUT-NOISE-TIMIT corpus for the evaluation of voice activity detection algorithms,” in *Interspeech*, Makuhari, Japan, 2010, pp. 26-30.

[26] A. Hyvärinen, ”Fast and robust fixed-point algorithms for independent component analysis,” *IEEE Trans. Neural Netw.*, vol. 10, pp. 626-634, 1999.

[27] A. Cichocki and S. Amari, ”Adaptive blind signal and image processing: learning algorithms and applications”, *John Wiley and Sons*, 2002.

[28] E. Vincent, R. Gribonval, and C. Fvotte, ”Performance measurement in blind audio source separation,” *IEEE Trans. Audio, Speech, Language Process.* , vol. 14, pp. 1462-1469, 2006.

[29] A. K. H. Al-Ali, D. Dean, B. Senadji, V. Chandran, and G. R. Naik, ”Enhanced forensic speaker verification using a combination of DWT and MFCC feature warping in the presence of noise and reverberation conditions,” *IEEE Access*, vol. 5, no. 99, pp. 15400-15413, 2017.

[30] S. O. Sadjadi, M. Slaney and L. Heck, ”MSR identity toolbox - A Matlab toolbox for speaker recognition research”, *Speech Language Process. Technical Committee Newslett.*, 2013.

[31] W.-C. Chen, C.-T. Hsieh and E. Lai, ”Multiband approach to robust text-independent speaker identification,” *J. Computational Linguistics Chinese Language Process.*, vol. 9, pp. 63-76, 2004.