Indonesian Journal of Electrical Engineering and Informatics (IJEEI)
Vol. 13, No. 3, September 2025, pp. 814∼829
ISSN: 2089-3272, DOI: 10.52549/ijeei.v13i3.5867

The Effect of Noise on Speaker Identification and Finding a Noise that Improves Accuracy

Md Atiqul Islam[1], Mohammed Abdul Kader[2]
[1]International Centre for Neuromorphic Systems, The MARCS Institute for Brain, Behaviour, and Development, Western Sydney University, Kingswood, NSW 2751, Australia
[2]Department of Electrical and Electronic Engineering, International Islamic University Chittagong, Bangladesh

Article Info

Article history:

Received September 4, 2024
Revised September 3, 2025
Accepted September 28, 2025

Keywords:

CAR-FAC
GFCC
GMM-UBM
Speaker Identification
Cocktail Party
Noise-robust

ABSTRACT

Conventional Speaker Identification (SID) systems accurately identify speakers if their speech is noiseless. However, their classification accuracies reduce substantially when speech is corrupted by noise. SID systems would be more practical and applicable if they were more noise-robust. We introduce an SID system that can accurately classify speakers, even when their speech is corrupted by various types of noise at different noise levels. We investigate the impact of noisy training data on the performance of an SID system and the noise that may enhance the performance of an SID system. In this paper, we compare two front-end feature extractors: a cochlea model called the Cascade of Asymmetric Resonators with Fast Acting Compression (CAR-FAC) and an FFT-based Gammatone Frequency Cepstral Coefficient (GFCC). We use the Gaussian Mixture Model with the Universal Background Model (GMM-UBM) and a Extreme Learning Machine (ELM) as classifiers to focus on the influence of the front-ends on performance. We train the GMM-UBM and the neural network with noisy data under various conditions to investigate the impact of noise on the classifier. Our results suggest that noisy training data make an SID system noise-robust while the performance under clean conditions remains almost the same. More interestingly, training with speech-shaped noise (cocktail party) enhances SID accuracy more than white noise.

Corresponding Author:

Mohammed Abdul Kader
Department of Electrical and Electronic Engineering
International Islamic University Chittagong
Chittagong, Bangladesh
Email: kader05cuet@gmail.com

1. INTRODUCTION

The applications of SID systems [1] are growing due to the advancement of the human-machine interface in state-of-the-art technologies. They find applications in a variety of online systems, including security systems [2], and call centers in banks [3] and health services [4]. Moreover, in the state-of-the-art, all Tesla's autopilot cars, Waymo's driverless cars and semi trucks [5], and all automated vehicles [6, 7] are becoming popular and part of our daily lives. Drivers can control these cars through a remote access voice control system in the car. Authentication of the driver's voice can secure access and the security of the vehicle. In addition, most smartphones include a voice authentication system to access their devices and data [8]. Moreover, banks such as HSBC and First Direct use SID systems for online and phone account customers.

Conventional SID methods apply FFT as a frequency analyzer. Mel-frequency Cepstral Coefficients (MFCC) [9, 10], Gammatone Frequency Cepstral Coefficients (GFCC) [11], and Power Normalized Cepstral

Coefficient (PNCC) [12] are examples of FFT-based methods and often achieve 100% accuracy when speech is clean, i.e., noiseless. In real-world applications, almost all input speech has some degree of background noise. The accuracy of FFT considerably methods drops when speech is corrupted by noise [13–16]. The FFT spectrum not only shows frequency harmonic but also shows energies related to frequencies. The addition of noise to a clean signal adds energy to the spectrum at varying frequencies. This causes a significant mismatch between a clean and noisy FFT spectrum and provides poor performance under noisy conditions.

In contrast, the human auditory system is famously robust to background noises and competing speeches, with the well-known cocktail-party problem as a typical example. Many papers have investigated the effect of cocktail party type noise on speech recognition performance [16–19]. Unfortunately, very few papers can be found showing the effect of a cocktail party on SID. In this work, we investigate the effect of cocktail party noise under training and testing conditions on a SID system.

Many cochlear models are now available that emulate auditory functions in humans. One particularly interesting model of the human cochlea is the Cascaded Asymmetric Resonators with Fast Acting Compression (CAR-FAC) model introduced by Lyon [20]. It has been shown to fit human physiological data better than the other six auditory models in response to relevant stimuli [12]. Our previous study [21] showed that the CAR-FAC and other cochlear methods achieve poor performance in fluctuating noise conditions. Thus, it is also a challenge for cochlear methods to provide noise-robust performance under noisy conditions.

There are two techniques to improve the performance of these methods under noisy conditions. The first technique focuses on altering the front-end feature extraction procedure to achieve noise-robust performance. Examples include speech enhancement techniques [22–24] and feature fusion [25–27]. The second technique is to train the classifier model in a way that will learn both types of noise and signal to enhance the SID performance. Here we investigate the second technique. The classifier speaker model also influences SID classification accuracy on noisy speech. Optimization of a classifier's parameters can strongly influence SID classification rates [28] . State-of-the-art back-ends, such as x vectors with deep neural networks [29] can perform very well in SID tasks with the cost of high computational time [30, 31]. The Gaussian Mixture Model with the Universal Background Model (GMM UBM) remains a popular classifier for SID systems due to its simple and fast implementation and strong performance on noiseless speech [8, 12, 13]. Like x-vectors, the GMM is also integrated with neural networks [32] to achieve better accuracy. In some cases, the GMM with the UBM outperforms neural networks [33] in a speaker identification task. It offers competitive performance to more recent classifiers, e.g., the i-vector, given clean or noisy speech [5, 11] with low computational cost. Moreover, the GMM-UBM can be a useful classifier to investigate changes in the front-end [21] as it does not add additional nonlinearities like neural networks.Nonlinear neural networks produce comparatively better performance than GMM-UBM. However, it requires larger training data to train them and the contribution of front-end data may not be understood correctly due to the nonlinearities in their mechanism. Like all classifiers, its performance degrades as the signal-to-noise ratio (SNR) of input speech decreases [3, 9]. This degradation is amplified when it is trained on clean speech, but the testing dataset is noisy. In this paper, we use a single layer neural network - the Extreme Learning Machine (ELM) [34], as used in the previous study [21] to examine if the same experimental setup could produce a similar result to the GMM-UBM and validate the integrity of this study. We did not explore a deep neural network due to limitations of data in our used datasets. The GMM-UBM exhibits poor performance in cases with nonlinear patterns of input data and mismatched conditions. The presence of noise in the input training data may reduce the discrepancy among speaker models under different levels of noise conditions and can play a significant role in producing noise-robust performance. Unlike the GMM-UBM, a neural network classifier requires large amounts of data and balanced data for training purposes to achieve better performance. The reason behind this type of training is to find the speaker's distinguishing features under a large variation of training data. We believe that noise in the training data may help to identify a speaker correctly with limited training data. This is one of the objectives in this paper.

This paper investigates if we can limit the performance degradation by training the GMM-UBM and ELM classifiers on noisy speech. We use the GFCC as an FFT algorithm and CAR-FAC as a cochlear algorithm to cover both types of auditory features. We investigate if the proposed technique to train the GMM-UBM, and ELM provides better performance for both algorithms. Moreover, a comparison of their performances is also presented here for the first time, using the YOHO dataset [35]. A recent study [36] has shown (in a mouse) that adding white noise to a signal can enhance the brain's ability to distinguish subtle tones by suppressing cortical tuning curves. In this study, we also investigate if the white noise-conditioned data in the speaker training is useful to achieve a noise-robust SID performance. We apply four different setups for UBM and

GMM modeling. Our results show that if we train our classifier with noisy data, we can achieve SID accuracy that is remarkably robust to noise.

2. METHODOLOGY

Figure 1 illustrates the block diagram of the presented SID system applying the CAR-FAC model. It is divided into training (top, Figure 1) and testing (bottom, Figure 1) stages. We describe each part in the following sections.

2.1. The front-end feature extraction procedure

The CAR-FAC model is described in [20]. Briefly, it uses a cascade of second-order asymmetric resonators to generate the Basilar Membrane (BM) response to a transduced traveling wave. The resonator pole and zero locations control the damping factor, which in turn changes the BM filter gain and bandwidth. They are responsible for the model's level dependent compressive nonlinearity [20]. The distance between the pole and zero, h, is then a crucial parameter in the CAR-FAC model.

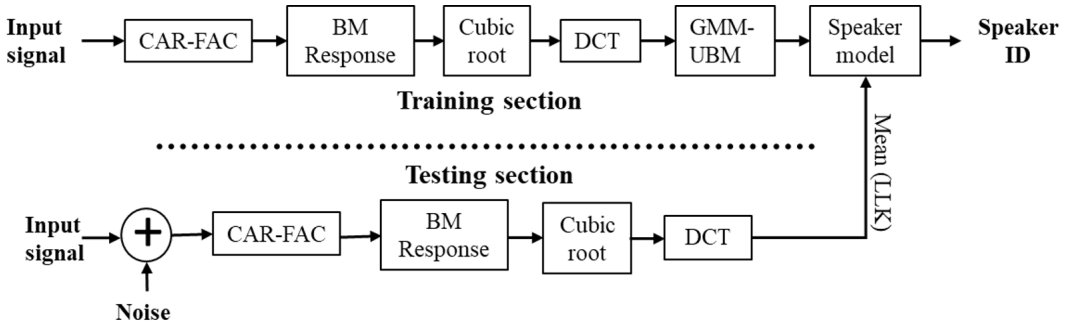

Figure 1. An illustration of the CAR-FAC SID system using the GMM-UBM, including training and testing phases. No samples are common in both training and testing datasets. In some experiments, we added noises with the training samples to generate noisy training data, as we detail later.

We set $h = sin(\frac{2\pi f_c}{f_s})$ to keep the pole a half-octave away from the zero location in the CAR-FAC model. Here, f_c is the characteristic frequency (CF) of each section of the cascade and f_s is the sampling frequency of the input signal. f_c is determined by the Greenwood function [37] to map 30 channels from 125 Hz to 3 kHz. We set the upper-frequency limit at 3 kHz because most SID cues, such as the speaker's fundamental frequency, pitch, and formants (f1 and f2) are below this cap [38]. In our investigation, we found the CAR-FAC with 70 channels produces a similar performance to 30 channels, but significantly slows down the back-end processing. Thus, we forward with the CAR-FAC with 30 channels.

The CAR-FAC algorithm implements nonlinear computations by controlling the pole (zero) radius r:

$$r = r_1 + d_{rz} \times (1 - b) \times NLF(v) \tag{1}$$

Here, r_1 is the minimum radius that maximally dampens the resonator:

$$r_1 = 1 - \xi \times \frac{2\pi f_c}{f_s} \tag{2}$$

where ξ is the damping factor and $d_{rz} = 0.7 \times (1 - r_1)$. In the CAR-FAC model, the damping factor controls the BM response compression. In human hearing research, typical damping factor values range from 0.1 to 0.4 [20]. We set the damping factor to 0.35.The level-dependent multi-rate nonlinearity comes from the inner hair cell feedback (b) through the Automatic Gain Control (AGC) loop filter. The instantaneous Non-Linear Function (NLF) interacts with the input waveforms velocity (v) in the CAR section of CAR-FAC.

$$NLF(v) = \frac{1}{1 + (kv + v_{offset})^2} \tag{3}$$

The NLF controls the gain of the CAR velocity and produces a combination tone, such as the cube-distortion tone in the cochlea [10].The parameters $k = 0.1$ and $v_{offset} = 0.04$ are default values in the

CAR-FAC implementation [20]. We apply these nonlinearities to obtain cochlear features. The BM energies are then computed as:

$$E(i) = \sum_{j=1}^{L} C(i, 1+j : j+L)^2 \tag{4}$$

where j is the starting index for each time window, L is the window duration, and $C(i)$ contains the output samples of channel i.

Next, we apply the cube root and DCT on the nonlinear BM energy feature to nonlinearly scale the features and decorrelate their dimensions. Moreover, the cube root and DCT application on the BM produces a better result than the only BM, as found in the previous study [21]. Figure 2 illustrates the effect of these transforms on BM energy in the form of a histogram. Cochlear features are not Gaussian distributed, as shown in Figure 2 (left panel), and features' data are highly correlated, as one bin covers most of the data range. Applying the cube root on the CAR-FAC features increases data variance, as shown in Figure 2 (middle panel). The cube root nonlinearly amplifies data to reduce the noise effect. This amplification hinders SID performance, as we observed empirically. Applying the DCT to the cube root features fits the data closer to a Gaussian distribution (right panel), which should help the GMM-UBM classifier build a better speaker model. We extract the GFCC following the previous study [11]. We use 64 channels for a frequency range of 125Hz to 3kHz. The Gammatone spectrum was down sampled to 100Hz, and only the magnitude spectrum was considered. Then, the cube root and DCT were applied. We omitted the first channel (which contains maximum energy and increases similarities among speakers). The channels above the 29th channel contain negligible energy compared to lower channels, and thus we omitted those channels. Therefore, the applied number of channels of the presented GFCC is 28.

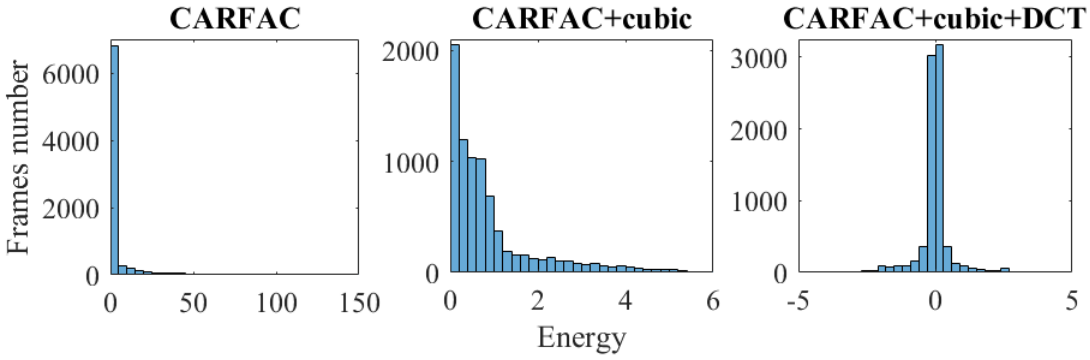

Figure 2. Histograms of the energy output of CAR-FAC, and the effect of the cube root and DCT on that output.

2.2. GMM-UBM speaker modeling

The GMM-UBM speaker modeling system has two parts: UBM development and an adaptation of the speaker data with the UBM to create a GMM speaker model. The UBM is a single GMM speaker model trained with all pooling data from the training or development dataset using the Expectation-Maximization (EM) algorithm [39]. The EM algorithm iteratively increases the likelihood of the dataset given GMM parameter values (weights, means, and variances). The estimation of the GMM parameters using the EM process has been detailed elsewhere [40].

In the GMM, each Gaussian component density $p_k(x)$ of a mixture component, k for input x is expressed as a function of a mean vector (μ_k) and a $D \times D$ variance matrix $(\sum_k)$ with a feature dimension D:

$$p_k(x) = \frac{1}{(2\pi)^{\frac{D}{2}} |\sum_k|^{\frac{1}{2}}} exp\{-\frac{1}{2}(x-\mu_x)^T \sum_k^{-1} (x-\mu_k)\} \tag{5}$$

The covariance matrix determines the correlation between adjacent feature dimensions. A diagonal covariance matrix is often used in the GMM instead of a full covariance matrix to restrict the Gaussian ellipse axis in the

direction of the coordinate axis [41]. This restriction helps the GMM to estimate better parameter values while requiring fewer samples and less computational time to do so. The parameter estimation of a full covariance GMM can be found in [42]. Note that the diagonal elements of a covariance matrix are the variances (σ^2) of channel features.

The UBM (λ_{UBM}) is defined by optimized weight W_k, mean μ_{1k}, and variance σ_{1k}^2 for all mixture components M as,

$$\lambda_{UBM} = \{W_k, \mu_{1k}, \sigma_{1k}^2\}, k = 1, 2, 3, \dots M \tag{6}$$

We used $M = 256$ mixtures in all experiments. The adaptation of GMM with the λ_{UBM} starts with all training samples $(x_t|t = 1, 2, 3 \dots T)$, and the probability $p(k|x_t)$ that a sample belongs to a particular mixture component:

$$p(k|x_t) = \frac{W_k p_k(x_t)}{\sum_{k=1}^{M} W_k p_k(x_t)} \tag{7}$$

$p(k|x_t)$ and x_t are used to compute the mixture probability counts (n), mean μ_2, and variance σ_2^2 for each mixture component using equation (7-9). This step is the same as the expectation step in UBM development.

$$n_k = \sum_{t=1}^{T} p(k|x_t) \tag{8}$$

$$\mu_{2k} = \frac{1}{n_k} \sum_{t=1}^{T} p(k|x_t) x_t \tag{9}$$

$$\sigma_{2k}^2 = \frac{1}{n_k} \sum_{t=1}^{T} p(k|x_t) x_t^2 \tag{10}$$

Here, n_k is a vector of length M, and the variance and mean sizes are $N \times M$, where N is the dimension of a feature. Finally, these new estimations from each speaker's training data are used to update the old statistics of the λ_{UBM} to create the adapted parameters for the GMM model for each mixture k using the following equations:

$$W_k^N = [\frac{(\alpha n_k)}{T} + (1 - \alpha) W_k]\gamma \tag{11}$$

$$\mu_k^N = \alpha E_k(x) + (1 - \alpha)\mu_k \tag{12}$$

$$\sum_{k}^{N} = \alpha E_k(x^2) + (1 - \alpha)(\sigma_{2k}^2 + \mu_k^2) - \mu_k^{N^2} \tag{13}$$

Here, α is the adaptation coefficient for the weights, means, and variances, respectively. This coefficient is essentially a learning rate, defined as a function of counts and a relevance factor (r), as $\alpha = \frac{n_k}{n_k + r}$. The γ is a scaling factor that ensures $\sum_{k=1}^{M} W_k = 1$.

In the testing stage, the log-likelihood of the vector sequence of a testing sample (X) is computed against each speaker model, and the mean of testing scores is computed using equation (12). Thus, each testing sample has a score against each speaker model.

$$X1 = \frac{\sum_{l=1}^{F} log(p(X_l|\lambda_{GMM}))}{F} \tag{14}$$

Here, $X1$ is a testing score against a speaker model for a target sample and F is the frame number. The maximum score for a testing sample against a speaker model indicates the target speaker identity.

2.3. Extreme Learning Machine (ELM) speaker modeling

We use the ELM instead of a deep neural network as it requires less training data for a comparatively better result. We use a single Long Short Term Memory (LSTM) [43] layer fully connected network with 400 neurons. This layer takes input as a sequence and produces series network equal to the number of speakers. This layer decides what information to keep, add, or discard and captures long-term features in input data. We use the Root Mean Square Propagation (RMSP) [44] optimization technique to train our network, with an initial learning rate of 0.001 and a regularization rate of 0.000005. The RMSP optimizer in ELM updates the output weights iteratively through modifying learning rate to speed up convergence. Moreover, the RMSP in ELM makes the training more stable. Interested reader can find details of RMSP in [44]. We resized our input feature to 64 × 64 to facilitate the ELM training. The maximum number of epochs for training was 22, and a batch size of 28 was used throughout. We used the SoftMax activation technique in the output layer. We trained the network with all types of noise at all SNR levels. Thus, there are 29 (4 SNRs × 7 types of noise + clean) speaker models. We used clean and noisy speaker models to identify the target speaker under all types of noise conditions. In the testing stage, the testing sample is compared with each speaker network to predict the speaker network to which the testing sample belongs. A testing sample produces only one output related to its class.

2.4. Dataset and Experimental Setup

We use the YOHO dataset [35] that previous SID systems employed [40, 45]. This dataset is relatively clean, which allows us to add noise according to the demand of our investigations. Which is why we have not used more realistic datasets, such as Voxceleb [46]. The YOHO dataset contains 138 speakers with 24 digit-pairs samples for each. We use 18 samples from each speaker to train the GMM and 1380 samples (10 from each speaker' training samples) to estimate the UBM parameters. The remaining 6 samples from each speaker were used for the testing purpose.

In many experiments, we apply noisy datasets to train the GMM and UBM. We add white, pink, street, restaurant, train, and car noises to clean signals to create noise-corrupted signals. The white noise and pink noise were generated using MATLAB, and the rest types of noise were downloaded from the $https : //www.freesfx.co.uk/$ website. The SNR ranges from 0 dB to 15 dB in steps of 5 dB. There are four types of training in this work. We use four training conditions:

i Clean GMM and clean UBM: We use clean samples from each speaker to train the GMM and the UBM.

ii Clean GMM and noisy UBM: We use clean 18 samples to train the GMM and noisy data to train the UBM. We use 306 samples from 17 speakers (8 speakers from 5dB and 9 speakers from 0dB) for each noise type. Thus, 119 speakers for seven different types of noise and the rest of 19 speakers with clean samples were used to train the UBM. This way, the UBM will learn the variability of noise types and levels. Eventually, the developed UBM will help the GMM to produce better SID accuracy under noisy and clean conditions.

iii Noisy GMM and clean UBM: We train the GMM using all types of noise with all SNRs. Thus, there are 28 (4 SNRs X 7 types of noise) GMM speaker models for noisy data and a single clean GMM speaker model for each speaker. In total, there are 29 speaker models for each speaker. In the real environment, it is unknown what type of noise is coming with the input signal, and our brain is trained on all possible combinations of noisy data. This is why we use all GMM speaker models to present all patterns of noise at probable SNRs. The test sample was tested against each speaker model, and the maximum matching score indicated the correct SID.

iv Noisy GMM and noisy UBM: We integrate the noisy UBM from experiment number ii and noisy GMM from experiment number iii to investigate if the presented training is beneficial to achieve a noise-robust SID score.

We use clean and noisy samples to test each developed speaker model. The speaker model that produces the maximum likelihood score of a testing sample from a speaker is considered the ID for that speaker.

3. RESULTS

The results are shown for the CAR-FAC and the GFCC method. Each figure presents the result showing the change when the GMM, UBM or both were trained using clean or noisy data. Each bar presents the average result of six random trials of training and testing samples. The error bar presents the minimum and maximum values of six trials. In the end, we also show the result using a single layer neural network to verify our findings using the GMM-UBM. A similar result with both classifiers will emphasize the noise in the training could improve SID accuracy.

3.1. Cascade of Asymmetric Resonators with Fast Acting Compression (CAR-FAC)

Figure 3 shows the SID result for the CAR-FAC method. The GMM-UBM was trained in clean and noisy conditions and tested under clean and seven different noise conditions. In clean or noisy training conditions, the CAR-FAC achieves almost 100% correct SID, as shown in Figure 3 (top, most right panel). However, the SID accuracy substantially drops with the increment of noise level, as shown in Figure 3 at 0 dB SNR for all types of noise.

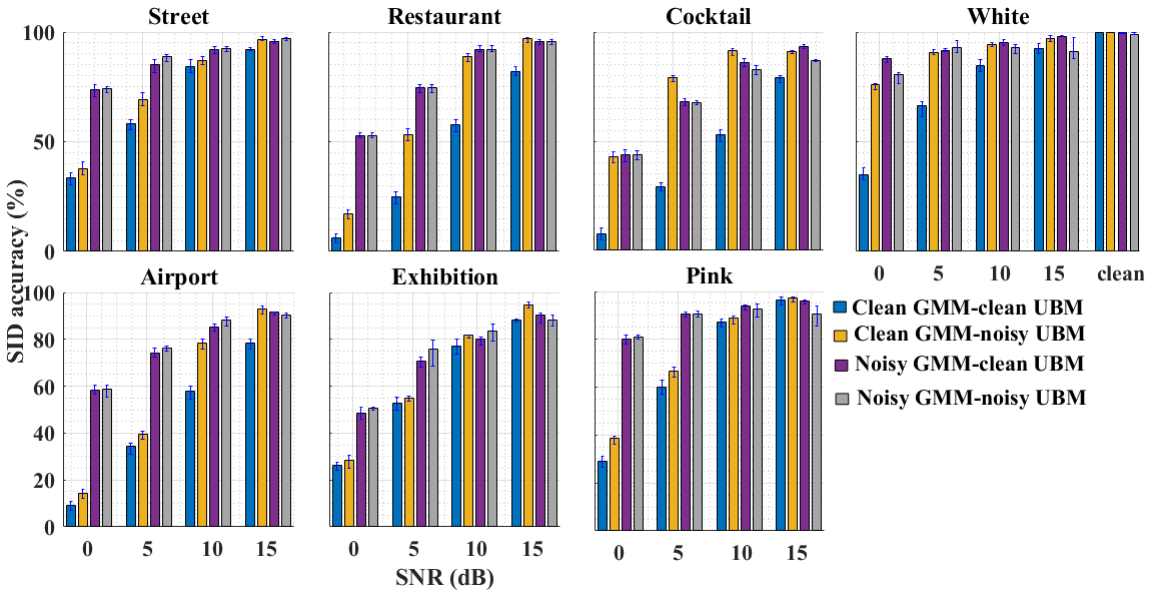

Figure 3. The SID results are shown for the CAR-FAC method under various types and levels of noise conditions. The GMM and UBM were trained with several conditioned of data. The legend shows the training conditions of the GMM and UBM. The title of each subplot indicates the noise used in the testing data while scoring a speaker.

To reduce the discrepancy in SID accuracy between low SNR and high SNR, we train the UBM with noisy data, while keeping the GMM clean. Figure 3 shows the presented training technique improves the SID accuracy at all SNR conditions for all types of noise while the SID accuracy in clean remained the same. This improvement is substantial, and the average SID improvement varies from 3.72% (Airport) to 34.04 % (cocktail), as shown in Figure 3. The UBM estimates the variation of noise and speaker variability. Thus, a noisy UBM tunes the parameters of the GMM to a better estimate of a speaker. Consequently, the developed GMM speaker model produces a high SID score, as shown in Figure 3. However, the decay rate of SID accuracy from clean to low SNR conditions is still almost linear. Next, we train the GMM using noisy data while keeping the UBM clean to investigate if the SID accuracy is further enhanced. The UBM is now only learning the speaker variability and adopting noisy GMM. Indeed, this technique further improves the SID accuracy for all types of noise under noisy conditions except for the cocktail noise. The cocktail noise is a heavily talked background noise that corrupts the speaker's speech and influences the GMM modelling. Which is why the GMM is producing poor SID accuracy under cocktail noise. We also train the GMM with a mixture of noise types and levels to observe which technique is more effective in generating a noise-robust SID. This new GMM with mixed noise gives poorer performance than our proposed training technique for individual GMM speaker models. The mixed noise raises higher variabilities among samples within a speaker. Hence,

the GMM estimates poor parameters from the speaker model. In the comparison of the noisy GMM and noisy UBM, the noisy UBM is way more effective than the noisy GMM to produce a noise-robust SID result.

Then, we train the GMM and the UBM with noisy data considering if this technique further improves the SID accuracy. Interestingly, this technique improves SID accuracy a little for some types of noise compared to the previous technique, as shown in Figure 3. For some types of noise, the SID accuracy is poorer than the previous technique. The noisy UBM meant to learn the speaker variability, and this can be achieved though clean speech training. The GMM is already noisy and the noisy UBM is poorly estimating speaker variability parameters. As a result, the noisy GMM and noisy UBM combination provides a poor SID score compared to the previous technique, as shown in Figure 3.

3.2. Gammatone Frequency Cepstral Coefficient (GFCC)

Figure 4 presents the result using the GFCC method. The same experimental setup following the experimental setup for the CAR-FAC was used for the GFCC method. The FFT-based GFCC produces 100% correct SID accuracy when the GMM-UBM is trained with clean or noisy data, as shown in Figure 4 (first bar in each subplot). However, this score reduces to almost 0% under mismatched conditions, except for the street noise. The energy distribution between clean and noisy FFT spectra makes them mismatched and causes havoc reduction of SID accuracy under noisy conditions. This reduction of performance is particularly true when the GMM-UBM was trained with clean data.

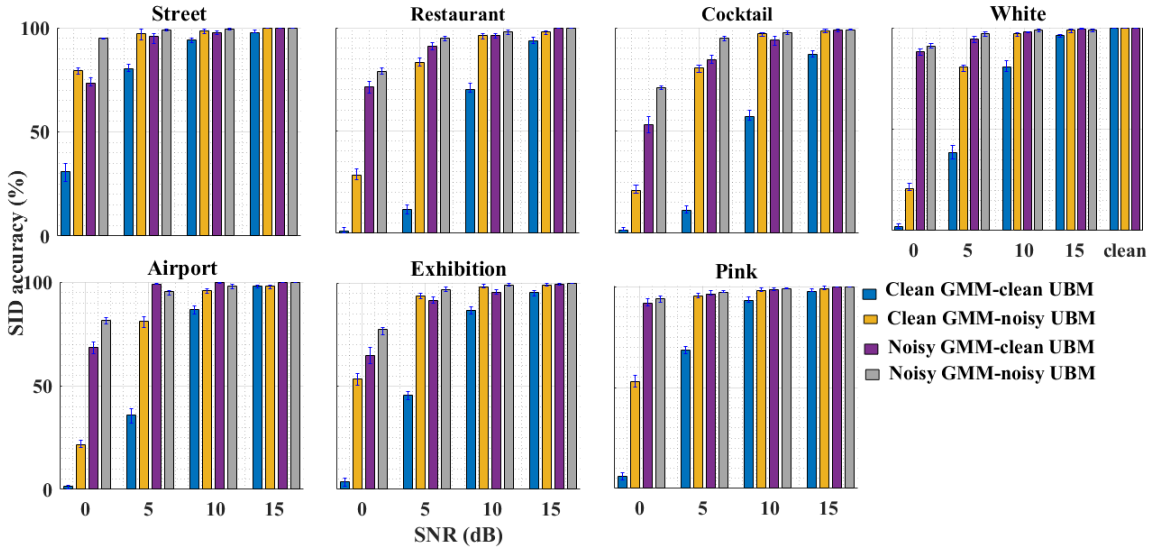

Figure 4. The SID results are shown for the GFCC method under various types and levels of noise conditions. The GMM and UBM were trained with several conditioned of data. The legend shows the training conditions of the GMM and UBM. The title of each subplot indicates the noise used in the testing data while scoring a speaker.

We introduce noise to UBM by training it with noisy data, while the GMM is trained with clean data. Figure 4 (second bar in each subplot) shows that the noisy UBM technique significantly enhances the SID accuracy. The presented method achieves an average 90% correct SID accuracy at 5 dB, as shown in Figure 4. Note that the 5 dB SNR is the noise threshold for an understandable conversation [47]. Comparing the performance of the clean UBM and noisy UBM, it can be claimed that the enhanced performance under noisy conditions comes from the noisy UBM. This enhancement evidence that the UBM not only learns the speaker variation but also models the noise variability while trained using noisy data. However, the SID score still needs to be improved under noisy conditions to claim the SID system is noise-robust. Next, we train the GMM using noisy data and the UBM with clean data considering if it further improves SID accuracy. Figure 4 (third bar in each subplot) shows the result using this technique. The presented technique substantially the performance of the GFCC method, except for the street noise. This improved result indicates that the noisy GMM training technique is more beneficial than the noisy UBM to produce a noise-robust SID accuracy.

The GFCC method further improves SID accuracy when both the GMM and the UBM were trained with noisy data. Figure 4 (fourth bar in each subplot) shows the generated results for this technique. Figure 4 shows that the performance of the presented technique produces almost consistent performance up to as low as 5 dB SNR irrespective of noise types. The average SID score at 0 dB SNR is more than 80%, as shown in Figure 4. Results in Figure 4 show that the noisy GMM and noisy UBM can learn both the speaker and noise variabilities and hence produce a noise-robust SID accuracy irrespective of types and levels of noise.

Comparing Figure 3 and Figure 4, on average, the CAR-FAC method provides a relatively better result than the GFCC method at lower SNR when the GMM-UBM was trained using clean data. This better result of the cochlear model under noisy conditions is also coherent with previous studies [21, 45, 48] [19, 39, 42]. This improvement is particularly true when the GMM-UBM is trained with clean data. The GFCC method has better results than the CAR-FAC method at higher SNRs. This better result is expected due to the matching of the energy spectrum of clean and high SNR data. The GFCC method has a better improvement than the CAR-FAC method while a noisy UBM is adopted with the clean GMM, as seen comparing Figure 3 and Figure 4. The CAR-FAC method has only an improved result than the GFCC method under white noise at 0 dB and 5 dB conditions. In contrast, the GFCC method produces significantly better performance over the CAR-FAC for all other types of noise at all SNR conditions.

Figure 3 and Figure 4 also show that the CAR-FAC method produces much improved SID accuracy over the GFCC method, while the GMM is trained with noisy data (third bar in each subplot). In contrast, the GFCC method produces a much-improved performance compared to the CAR-FAC method, while a noisy UBM is adopted with the clean or noisy GMM. This improvement is observed in Figure 3 and 4 (the third and fourth bars in each subplot). The utilization of noisy data in the GMM-UBM training enhances the performance of the FFT-based GFCC method more than the cochlear-based CAR-FAC method. Both methods exhibit the poorest performance in the cocktail party noise scenario. The cocktail party noise is challenging as it contains many voices from party participants. However, both methods produce an improved performance under noisy training of the GMM-UBM. Thus, the exposure of noisy data at the classifier is necessary to develop a noise-robust SID system.

3.3. Noise that improves SID performance

We use the GMM-UBM and a single-layer neural network to investigate the noise in the training of an SID system that enhances SID accuracy. In the following section, we present results for those two classifiers.

3.3.1. Results using the GMM-UBM

The study [36] found that the use of white noise in the brain of a mouse improves the brain's ability to separate subtle tone discrepancies by suppressing cortical tuning curves. They claimed that white noise is a good noise that enhances auditory perception. Moreover, we observed in the previous experiments in Figures 3 to 4 that the performance for both methods struggles under cocktail noise. Thus, it is interesting to investigate which noise in training or testing produces a better SID result. In machine learning, the classifier can be considered as the brain that distinguishes speakers based on their speeches. The presence of noise in the classifier also enhances SID accuracy performance, as observed in Figure 3 to Figure 4. Here, we investigate, "Which noise in the training data enhances the performance of a classifier?". The advantage of using the GMM-UBM is that it can focus only on the changes in front- ends without adding nonlinearities like a neural network. We have trained the GMM using each type of noise-conditioned data and tested it under all types of noisy data. Note that the UBM was trained on clean data.

We used the CAR-FAC and GFCC methods to investigate which noise-trained classifier produces the best SID accuracy. We use the same experimental setup for both methods. The investigated results are separately shown in Figure 5 and Figure 6 for the CAR-FAC and GFCC, respectively. Each subplot title shows the noise that was used to train the classifier. The legend shows the types of noise used to test the classifier. The average results at an individual SNR are also shown to determine the best result at a specific noise trained condition of the GMM-UBM classifier.

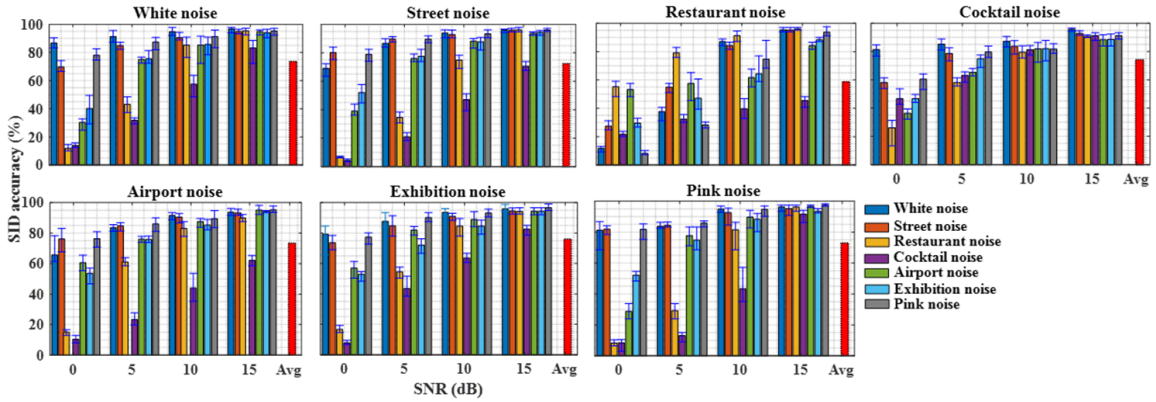

Figure 5. The presentation of the performance of the CAR-FAC method under different noise training and testing under other noise conditions. The results are shown for various SNR conditions. An average result for each noise-conditioned training is shown with a red bar (most right bar in each subplot). The errorbar presents the minimum and maximum values of six trials.

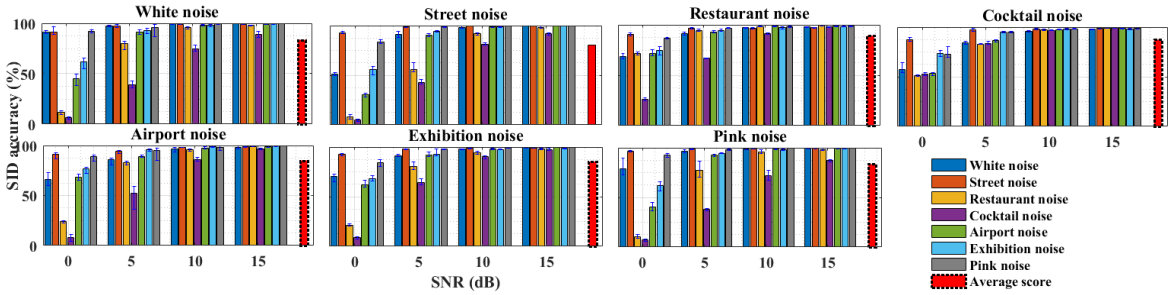

Figure 6. The presentation of the performance of the GFCC method under different noise training and testing under other noise conditions. The results are shown for various SNR conditions. An average result for each noise-conditioned training is shown with a red bar (most right bar in each subplot). The errorbar presents the minimum and maximum values of six trials.

The presence of individual noise in the speaker training enhances the performance of the classifier while the same type of noise is added to the testing signal. However, the addition of noise in the speaker training may enhance or reduce the performance for other types of noise. This scenery has been shown in Figures 5 and 6. The results are shown as an average for six trials of random selections of training and testing samples. The error bar indicates the minimum and maximum values of six trials. The CAR-FAC has the best average performance while the exhibition noise-conditioned data are used to train the classifier, as shown in Figure 5. However, this method provides relatively better performance for all testing noise conditions at 0 dB, while the cocktail noise was used in the speaker training. The average SID accuracy under the cocktail noise training condition is also similar to the exhibition noise training condition. It can be predicted that the presence of the speech-pattern noise type, such as the cocktail and exhibition noise, at the classifier is more beneficial than other types of noise, as shown in Figure 5. In contrast, these types of noise produce a poor performance while they were present in the testing data. This result indicates that the cocktail party noise is the most challenging while testing for a SID system to achieve a noise-robust performance. The performance of the GFCC method has been shown in Figure 6. The GFCC method provides similar performance at 15 dB SNR for all types of noise training conditions. However, the performance is subjective to training noise type at low SNR conditions, as shown in Figure 6. Unlike the result shown in Figure 5, the GFCC method produces a poor SID accuracy for the cocktail and restaurant noise while other types of noise are used to train the speaker classifier. Interestingly, this method produces a better result while the cocktail or restaurant noise is used to train the speaker classifier. Note that the restaurant noise is also a speech-shaped noise.

Comparing Figure 5 and Figure 6, the GFCC provides a better result than the CAR-FAC method. The deviation of minimum and maximum values for each testing condition are less varied for the GFCC method compared to the CAR-FAC method. Both methods provide poor performance while speech-shaped noise such as cocktail or restaurant noise was used for testing. This indicates that the speech-shaped noise is more challenging compared to other types of noise while they are added to testing signals. In contrast, the speech shaped noise is good while used for speaker training and provides better performance for other types of noise conditioned testing. Distinctive results from two different methods suggest that the front-end processor plays a vital role to process noisy data and contribute to the performance of a classifier. Eventually, the performance of the classifier significantly improves in the presence of noise in the signal.

The investigation suggests that the white noise may enhance SID accuracy under high SNR conditions. However, the effect of white noise may not be significant at low SNRs. In contrast, the speech-shaped noise, such as cocktail noise in the classifier may be more useful to produce a noise-robust performance under adverse conditions. Thus, the white noise based speaker training may not be good for a SID task, though this noise may enhance speech perception in the brain [36].

3.3.2. CAR-FAC and neural network-based result

We also a neural network to investigate if noise in the training of an SID system can improve SID accuracy. A deep neural network would require a much bigger dataset to properly tune parameters and achieve a state-of-the-art result. Hence, we used a simple fully connected neural network with one hidden layer. We use the same setup for the neural network as was used for the GMM-UBM, but we resized the input sample to 64 by 64 to facilitate and speed up the training of the network. Thus, we expect that our neural network will produce a poorer SID accuracy than the GMM-UBM for our used dataset. The generated results are shown in Figure 7. The average results shown in Figure 7 are lower than the results shown in Figure 5, which supports our assumption. Figure 7 shows a similar result to Figure 5 and 8. The presence of white noise during training produces the poorest performance during testing, as shown in Figure 7. The highest SID accuracy was observed under airport noise, which is 77% on average, as shown in Figure 7. Other types of noise produce similar average results, as shown in Figure 7. Again, training and testing with the same type of noise produces the highest SID accuracy for all types of noise.

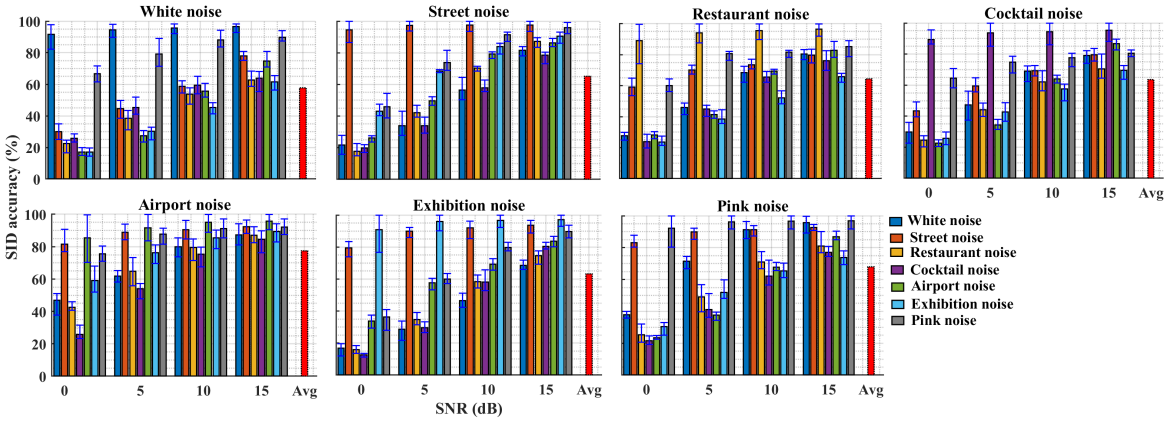

Figure 7. The presentation of the performance of the CAR-FAC with the neural network under different noise training and testing under other noise conditions. The results are shown for various SNR conditions. An average result for each noise-conditioned training is shown with a red bar (most right bar in each subplot). The error bar presents the minimum and maximum values of six trials.

3.3.3. Discussion

In this study, we investigate the impact of noise building a noise-robust SID system. We used the FFT-based GFCC and cochlear-based CAR-FAC methods for our investigation. The cochlear front-end produces much better noise-robust performance than the FFT method when the GMM-UBM is trained with clean samples. However, the SID accuracy decreased substantially under low SNR conditions. Utilization of noisy data to train either the GMM or the UBM improves SID accuracies under noisy test conditions.

The training of the UBM with noisy data enhances the performance of the FFT method more than the cochlear method. The CAR-FAC channel information is more correlated than the FFT-based GFCC's channel information, which may hinder the UBM to learn speaker variation to produce a better universal speaker model for the CAR-FAC compared to the GFCC. In contrast, the CAR-FAC method produces better performance than the GFCC method when the GMM is trained with noisy data. This improved result of the CAR-FAC indicates that the CAR-FAC extracts speaker distinguishing features, which are useful for the GMM to make individual speaker models. When we train the GMM-UBM on noisy data, it can more accurately classify noisy speech, while the accuracy in clean conditions remains intact.

Biological systems do not ignore the noise in their environments but instead seem to learn them [**?**]. Perhaps one reason that humans identify speakers from noisy speech so well is that they learn accurate noise models throughout their lives and apply them in suitable environments. We observed that the individual noise in the training samples is very effective to identify a target speaker under that type of noise condition. This performance is valid for all types of noise and makes the SID system consistent in producing noise-robust performance. We also tried to investigate if a mixture of all types of noise at different levels can influence the performance of an SID system. We randomly selected four speakers for each SNR (16 speakers for each noise) and seventeen speakers for clean conditions. Interestingly, we found that this combination of noise reduces SID accuracy significantly under clean and noisy conditions. This degradation suggests that the human auditory system may know noise types at each SNR level, which makes the auditory system noise-robust and consistent in performance with the SID task.

If we consider the human auditory system performing an SID task, then the cochlea is its front-end feature extractor, and the brain is its classifier. We do not fully understand how the brain helps humans disambiguate speech from noise. Our work does not shed light on this issue, because the GMM-UBM classifier is not a biologically inspired system. We applied it because the GMM-UBM is popular [30, 31, 33], simple, computationally fast, and fit for focusing only on front-end changes. Other classifiers, such as neural networks more closely resemble biological structure and function. We applied the neural network to present a biological SID system. The neural network produces poorer performance than the GMM-UBM. This could be due to less available training data in our dataset. However, the neural network produces a similar pattern of results to the GMM-UBM. Our results do indicate that learning noise during training enhances recognition performance and makes the system more robust to noise. This outcome supports the findings of previous studies [49–51]. Noise-robust performance using a noisy data-trained classifier indicates that noise might be one of the influencing factors on brain signals and recognition performance. Thus, we also investigate which type of noise enhances SID performance using seven different types of noise. The performance of the presented methods fluctuated with the variation of noise types. However, cocktail party noise is considered the most challenging under testing conditions for a SID system regardless of front-end processing. Many studies have been performed on speech recognition in a cocktail party environment, but unfortunately, very few on speaker identification. This work can be an example to show the effect of a cocktail party on a SID system.

The classifier trained with white noise data generates a better result for testing data with added stationary or slow-varying (pink or street) noise. The testing data with fluctuating types of noise (restaurant, exhibition, cocktail, or airport) produce poor performance when the classifier is trained with white noise data. This result was observed for both classifiers. In contrast, the fluctuating noise in speaker models enhances the performance and produces an improved SID accuracy for all other types of noise. The study in [36] showed that added white noise improves subtle tone discrimination based on a mouse. In contrast, we discovered that fluctuating (restaurant, cocktail, or airport) noise is better than white noise to distinguish speakers in an SID system. The outcome of this study can be extended for speech recognition, SID for cochlear implant patients, speech intelligibility, and sound localization to develop a noise-robust SID system. A bigger dataset with the cochlear front-end and deep neural network can be investigated to execute a more biologically plausible SID system. Finally, we need to apply noise to develop a speaker model to implement a noise-robust SID system. This technique can be an alternative to the denoising technique to achieve a noise-robust SID system.

4. CONCLUSION

In this study, we investigate the impact of noise building a noise-robust SID system. We used the FFT-based GFCC and cochlear-based CAR-FAC methods for our investigation. The cochlear front-end produces much better noise-robust performance than the FFT method when the GMM-UBM is trained with clean

samples. However, the SID accuracy decreased substantially under low SNR conditions. Utilization of noisy data to train either the GMM or the UBM improves SID accuracies under noisy test conditions. The outcome of this study can be extended for speech recognition, SID for cochlear implant patients, speech intelligibility, and sound localization to develop a noise-robust SID system. A bigger dataset with the cochlear front-end and deep neural network can be investigated to execute a more biologically plausible SID system. Finally, we need to apply noise to develop a speaker model to implement a noise-robust SID system. This technique can be an alternative to the denoising technique to achieve a noise-robust SID system.

REFERENCES

[1] B. Saritha, M. A. Laskar, and R. H. Laskar, *A Comprehensive Review on Speaker Recognition*. Cham: Springer International Publishing, 2023, pp. 3–23. [Online]. Available: https://doi.org/10.1007/978-3-031-18444-4_1

[2] N. Saxena and D. Varshney, "Smart home security solutions using facial authentication and speaker recognition through artificial neural networks," *International Journal of Cognitive Computing in Engineering*, vol. 2, pp. 154–164, 2021.

[3] L. Nosrati, A. M. Bidgoli, and H. H. S. Javadi, "Machine learning and metaheuristic algorithms for voice-based authentication: A mobile banking case study," *International Journal of Computational Intelligence Systems*, vol. 17, no. 1, p. 287, 2024.

[4] Y. Kumar, "A comprehensive analysis of speech recognition systems in healthcare: current research challenges and future prospects," *SN Computer Science*, vol. 5, no. 1, p. 137, 2024.

[5] J. M. Scanlon, K. D. Kusano, T. Daniel, C. Alderson, A. Ogle, and T. Victor, "Waymo simulated driving behavior in reconstructed fatal crashes within an autonomous vehicle operating domain," *Accident Analysis & Prevention*, vol. 163, p. 106454, 2021.

[6] C. Badue, R. Guidolini, R. V. Carneiro, P. Azevedo, V. B. Cardoso, A. Forechi, L. Jesus, R. Berriel, T. M. Paixao, F. Mutz *et al.*, "Self-driving cars: A survey," *Expert systems with applications*, vol. 165, p. 113816, 2021.

[7] M. Anjum and S. Shahab, "Improving autonomous vehicle controls and quality using natural language processing-based input recognition model," *Sustainability*, vol. 15, no. 7, p. 5749, 2023.

[8] F. Thullier, "A practical application of a text-independent speaker authentication system on mobile devices," Ph.D. dissertation, Université du Québec à Chicoutimi, 2016.

[9] D. P. Ellis, "Plp and rasta and mfcc, and inversion in matlab," Feb. 2024, http://surl.li/tozjk.

[10] S. Davis and P. Mermelstein, "Comparison of parametric representations for monosyllabic word recognition in continuously spoken sentences," *IEEE transactions on acoustics, speech, and signal processing*, vol. 28, no. 4, pp. 357–366, 1980.

[11] Y. Shao, S. Srinivasan, and D. Wang, "Incorporating auditory feature uncertainties in robust speaker identification," in *2007 IEEE International Conference on Acoustics, Speech and Signal Processing-ICASSP'07*, vol. 4. IEEE, 2007, pp. IV–277.

[12] Q. Li and Y. Huang, "An auditory-based feature extraction algorithm for robust speaker identification under mismatched conditions," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 6, pp. 1791–1801, 2010.

[13] M. S. Alam and M. S. Zilany, "Speaker identification system under noisy conditions," in *2019 5th International Conference on Advances in Electrical Engineering (ICAEE)*. IEEE, 2019, pp. 566–569.

[14] A. Ashar, M. S. Bhatti, and U. Mushtaq, "Speaker identification using a hybrid cnn-mfcc approach," in *2020 International Conference on Emerging Trends in Smart Technologies (ICETST)*. IEEE, 2020, pp. 1–4.

[15] M. Tiwari and D. K. Verma, "Enhanced text-independent speaker recognition using mfcc, bi-lstm, and cnn-based noise removal techniques," *International Journal of Speech Technology*, vol. 27, no. 4, pp. 1013–1026, 2024.

[16] G. M. Bidelman, F. Bernard, and K. Skubic, "Hearing in categories and speech perception at the "cocktail party"," *PloS one*, vol. 20, no. 1, p. e0318600, 2025.

[17] W. Zhang, C. Boeddeker, S. Watanabe, T. Nakatani, M. Delcroix, K. Kinoshita, T. Ochiai, N. Kamo, R. Haeb-Umbach, and Y. Qian, "End-to-end dereverberation, beamforming, and speech recognition with improved numerical stability and advanced frontend," in *ICASSP 2021-2021 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*. IEEE, 2021, pp. 6898–6902.

[18] L. Lu, Y. Ding, C. Xue, and L. Li, "Negative emotions in the target speaker's voice enhance speech recognition under "cocktail-party" environments," *Attention, Perception, & Psychophysics*, vol. 83, pp. 247–259, 2021.

[19] Y.-m. Qian, C. Weng, X.-k. Chang, S. Wang, and D. Yu, "Past review, current progress, and challenges ahead on the cocktail party problem," *Frontiers of Information Technology & Electronic Engineering*, vol. 19, pp. 40–63, 2018.

[20] R. F. Lyon, *Human and machine hearing: extracting meaning from sound*. Cambridge University Press, 2017.

[21] M. A. Islam, Y. Xu, T. Monk, S. Afshar, and A. van Schaik, "Noise-robust text-dependent speaker identification using cochlear models," *The Journal of the Acoustical Society of America*, vol. 151, no. 1, pp. 500–516, 2022.

[22] X. Zhao, Y. Shao, and D. Wang, "Casa-based robust speaker identification," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 20, no. 5, pp. 1608–1616, 2012.

[23] N. Wang, P. Ching, N. Zheng, and T. Lee, "Robust speaker recognition using denoised vocal source and vocal tract features," *IEEE transactions on audio, speech, and language processing*, vol. 19, no. 1, pp. 196–205, 2010.

[24] N. Wang, P. C. Ching, N. Zheng, and T. Lee, "Robust speaker recognition using denoised vocal source and vocal tract features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 19, no. 1, pp. 196–205, 2011.

[25] M. T. Al-Kaltakchi, W. L. Woo, S. S. Dlay, and J. A. Chambers, "Comparison of i-vector and gmm-ubm approaches to speaker identification with timit and nist 2008 databases in challenging environments," in *2017 25th European Signal Processing Conference (EUSIPCO)*. IEEE, 2017, pp. 533–537.

[26] R. Jahangir, Y. W. Teh, N. A. Memon, G. Mujtaba, M. Zareei, U. Ishtiaq, M. Z. Akhtar, and I. Ali, "Text-independent speaker identification through feature fusion and deep neural network," *IEEE Access*, vol. 8, pp. 32 187–32 202, 2020.

[27] V. Karthikeyan, S. S. Priyadharsini, and K. Balamurugan, "Attention-based multi dimension fused-feature convolutional neural network framework for speaker recognition," *Multimedia Tools and Applications*, pp. 1–31, 2025.

[28] Q. Hong, S. Kwong, and H. Wang, "Optimization of gaussian mixture model parameters for speaker identification," in *Genetic and Evolutionary Computation Conference*. Springer, 2004, pp. 1310–1311.

[29] D. Snyder, D. Garcia-Romero, G. Sell, D. Povey, and S. Khudanpur, "X-vectors: Robust dnn embeddings for speaker recognition," in *2018 IEEE international conference on acoustics, speech and signal processing (ICASSP)*. IEEE, 2018, pp. 5329–5333.

[30] D. A. Reynolds, T. F. Quatieri, and R. B. Dunn, "Speaker verification using adapted gaussian mixture models," *Digital signal processing*, vol. 10, no. 1-3, pp. 19–41, 2000.

[31] J. H. Hansen and T. Hasan, "Speaker recognition by machines and humans: A tutorial review," *IEEE Signal processing magazine*, vol. 32, no. 6, pp. 74–99, 2015.

[32] A. Bakshi and S. K. Kopparapu, "Spoken indian language classification using gmm supervectors and artificial neural networks," in *2019 IEEE Bombay Section Signature Conference (IBSSC)*. IEEE, 2019, pp. 1–6.

[33] M. Fasounaki, E. B. Yüce, S. Öncül, and G. İnce, "A comparative assessment of text-independent automatic speaker identification methods using limited data," *Avrupa Bilim ve Teknoloji Dergisi*, no. 26, pp. 217–222, 2021.

[34] G.-B. Huang, Q.-Y. Zhu, and C.-K. Siew, "Extreme learning machine: theory and applications," *Neurocomputing*, vol. 70, no. 1-3, pp. 489–501, 2006.

[35] J. Campbell and A. Higgins, "Yoho speaker verification," *Linguistic Data Consortium, Philadelphia*, 1994.

[36] R. K. Christensen, H. Lindén, M. Nakamura, and T. R. Barkat, "White noise background improves tone discrimination by suppressing cortical tuning curves," *Cell reports*, vol. 29, no. 7, pp. 2041–2053, 2019.

[37] D. D. Greenwood, "Critical bandwidth and the frequency coordinates of the basilar membrane," *The Journal of the Acoustical Society of America*, vol. 33, no. 10, pp. 1344–1356, 1961.

[38] J. C. Stemple, N. Roy, and B. K. Klaben, *Clinical voice pathology: Theory and management*. Plural Publishing, 2018.

[39] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society: series B (methodological)*, vol. 39, no. 1, pp. 1–22, 1977.

[40] D. A. Reynolds and R. C. Rose, "Robust text-independent speaker identification using gaussian mixture speaker models," *IEEE transactions on speech and audio processing*, vol. 3, no. 1, pp. 72–83, 1995.

[41] T. Kinnunen and H. Li, "An overview of text-independent speaker recognition: From features to supervectors," *Speech communication*, vol. 52, no. 1, pp. 12–40, 2010.

[42] K.-H. Yuo and H.-C. Wang, "Joint estimation of feature transformation parameters and gaussian mixture model for speaker identification," *Speech Communication*, vol. 28, no. 3, pp. 227–241, 1999.

[43] A. Graves and A. Graves, "Long short-term memory," *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

[44] A. Graves, "Long short-term memory," *Supervised sequence labelling with recurrent neural networks*, pp. 37–45, 2012.

[45] M. A. Islam, W. A. Jassim, N. S. Cheok, and M. S. A. Zilany, "A robust speaker identification system using the responses from a model of the auditory periphery," *PloS one*, vol. 11, no. 7, p. e0158520, 2016.

[46] A. Nagrani, J. S. Chung, and A. Zisserman, "Voxceleb: a large-scale speaker identification dataset," *arXiv preprint arXiv:1706.08612*, 2017.

[47] J. H. Rindel, "Restaurant acoustics–verbal communication in eating establishments," *Acoustics in practice*, vol. 7, no. 1-14, 2019.

[48] M. Islam, M. Zilany, and A. Wissam, "Neural-response-based text-dependent speaker identification under noisy conditions," in *International Conference for Innovation in Biomedical Engineering and Life Sciences: ICIBEL2015, 6-8 December 2015, Putrajaya, Malaysia 1*. Springer, 2016, pp. 11–14.

[49] L. Zhang, F. Schlaghecken, J. Harte, and K. L. Roberts, "The influence of the type of background noise on perceptual learning of speech in noise," *Frontiers in Neuroscience*, vol. 15, p. 646137, 2021.

[50] G. B. Söderlund, J. Åsberg Johnels, B. Rothén, E. Torstensson-Hultberg, A. Magnusson, and L. Fälth, "Sensory white noise improves reading skills and memory recall in children with reading disability," *Brain and Behavior*, vol. 11, no. 7, p. e02114, 2021.

[51] M. J. Jafari, R. Khosrowabadi, S. Khodakarim, and F. Mohammadian, "The effect of noise exposure on cognitive performance and brain activity patterns," *Open access Macedonian journal of medical sciences*, vol. 7, no. 17, p. 2924, 2019.

BIOGRAPHIES OF AUTHORS

Md Atiqul Islam Dr. Md Atiqul Islam has received his PhD from Western Sydney University. Currently, he is a research assistant at the same university and a casual lecturer at Kent Institute, Australia. He completed his MSc in Engineering from the University of Malaya, Malaysia. He finished his BSc in Electrical and Electronic Engineering from Rajshahi University of Engineering and Technology. He has teaching experience at different universities in various countries for more than eight years. His research interest covers hearing research, machine learning, natural language processing, renewable energy, and ethics and privacy in information technology. Email: atiqatrai@gmail.com. Please visit https://scholar.google.com.au/citations?user=uMa5IJAAAAAJ&hl=en to know more about his research.

Mohammed Abdul Kader is working as a faculty member in the Department of Electrical and Electronic Engineering at the International Islamic University, Chittagong, Bangladesh. He is also pursuing a part-time Ph.D. in Electrical and Electronic Engineering at Chittagong University of Engineering and Technology, Bangladesh, where he also earned his undergraduate and graduate degrees. He worked as a researcher from October 1, 2024, to July 31, 2025, in the Advanced Nuclear Technology Research Group in the Department of Nuclear Engineering at Universitat Politècnica de Catalunya, Barcelona, Spain. To date, he is the author or co-author of 43 Scopus-indexed research papers. His research interests include robotics, embedded systems, machine learning, and digital signal processing. He can be contacted at: kader05cuet@gmail.com.