Noise-Robust Speaker Recognition Based on Morphological Component Analysis

Yongjun He[1], *Chen Chen*[2], *Jiqing Han*[2]

[1]School of Computer Science and Technology, Harbin University of Science and Technology
[2]School of Computer Science and Technology, Harbin Institute of Technology, Harbin, China

heyongjun@hit.edu.cn, chenc53@126.com, jqhan@hit.edu.cn

Abstract

Speaker recognition suffers severe performance degradation under noisy environments. To solve this problem, we propose a novel method based on morphological component analysis. This method employs a universal background dictionary (UBD) to model common variability of all speakers, a speech dictionary of each speaker to model special variability of this speaker and a noise dictionary to model variability of environmental noise. These three dictionaries are concatenated to be a big dictionary, over which test speech is sparsely represented and classified. To improve the discriminability of speaker dictionaries, we optimize the speaker dictionaries by removing speaker atoms which are close to the UBD atoms. To ensure varying noises can be tracked, we design an algorithm to update the noise dictionary with the noisy speech. We finally conduct experiments under various noise conditions and the results show that the proposed method can obviously improve the robustness of speaker recognition under noisy environments.

Index Terms: Morphological component analysis, sparse representation, discriminant dictionary, speaker recognition

10.21437/Interspeech.2015-96

1. Introduction

Speaker recognition has drawn considerable attention in the past few decades. The state-of-the-art techniques, such as the GMM-UBM [1], GMM-SVM [2] and joint factor analysis [3], have achieved satisfied performances in ideal conditions or in session-varied conditions. However, their performances suffers severely degradation under noisy environments [4], which prevents these techniques from real applications.

Many methods can be used to enhance the noise robustness of speaker recognition. The first category is the feature normalization, including the cepstral mean normalization (CMN) [5], cepstral mean and variance normalization (CMVN) [6], histogram equalization [7], MVA post-processing[8], etc. These methods can only get limited improvement because they do not learn any information about noise. The second category is to remove noise from noisy speech by enhancement methods, such as the spectrum subtraction and Wiener filtering [9], and then extract feature from the enhanced speech. The problem is that, it is difficult to model and estimate the noise because it is often non-stationary and even speech-like. As a result, speech enhancement methods inevitable cause further distortion, to which current speaker recognition methods are sensitive. Therefore, we need new techniques to solve this problem.

In the past few years, sparse coding [10] has been extensively investigated and provides possible solutions to speaker recognition under noisy environments. This technique represents signals with a set of atoms (elemental signals) from a dictionary (atom collection). By sparse coding, we mean that the representation accounts for most or all information of a signal, with a linear combination of only a small number of atoms. Recently, one sparse coding method called morphological component analysis (MCA) [11] has been applied successfully to speaker recognition [12] [13] [14]. Based on this technique, one dictionary for each speaker is prepared and all speaker dictionaries are concatenated to be a big dictionary. In recognition, the test speech is represented sparsely over the big dictionary. In theory, the speech spoken by one speaker can only be represented by the dictionary of this speaker; therefore, the representation can be directly used for classification.

Almost all reported speaker recognition methods based on sparse coding employ the MCA framework. These methods first transform training speech into GMM mean supervectors [12] [13] or total variability i-vectors [14], and then combine these vectors to be a big dictionary, over which the sparse decomposition and classification are made. It is reported that these methods have achieved better performances than traditional GMM-UBM and GMM-SVM [12]. However, there are two problems needed to be addressed for a further improvement. First, the MCA can perform as expected only when its assumption holds true, i.e., the speech from one speaker is sparse over the dictionary of this speaker but is dense over the other dictionaries [11]. However, there is no proof in theory or in experiment to show that this assumption holds true in the GMM mean supervector domain or in the i-vector domain. Second, the noise robustness is still not considered in current MCA-based methods, resulting in bad performances under noisy environments.

To solve these problems, we propose a novel speaker identification method based on MCA. In this method, signals are transformed into the magnitude spectrum domain, where speech is prone to be sparse. Besides, dictionaries are prepared by a training way which can ensure a sparser result compared with exemplar collection. More importantly, the speaker dictionaries are further optimized by removing atoms which are close to the UBD atoms, leading to discriminant speaker dictionaries. Furthermore, a noise dictionary updated with the input noise is employed to ensure that time-varying noise can be absorbed.

This research is partly supported by the National Natural Science Foundation of China under grant No. 61305001, 91120303 and 91220301, the Scientific Research Fund of Heilongjiang Provincial Education Department under grant No. 12511096, the Natural Science Foundation of Heilongjiang Province under grant No. F200936.

2. MCA and speaker recognition

MCA is proposed to separate the mixture of signals having different morphologies, and it has been applied successfully in image separation. Given a signal $Y \in \mathbb{R}^N$ which is the mixture of K different signals $X_1, X_2, ..., X_K$, i.e.,

$$Y = \sum_{i=1}^{K} X_i, \tag{1}$$

there is a big dictionary $\mathbf{\Psi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_K]$ with each $\mathbf{\Phi}_i \in \mathbb{R}^{N \times M_i}$ being a dictionary of X_i, where K is the number of dictionaries and M_i is the atom number in $\mathbf{\Phi}_i$. MCA assumes that each signal X_i is sparse over $\mathbf{\Phi}_i$ but is dense over other $\mathbf{\Phi}_j$ (for all $j \neq i$). For the signal Y, MCA solves the following optimization problem

$$\min_{\mathbf{y}} \|\mathbf{y}\|_0 \text{ subject to } \|Y - \mathbf{\Psi}\mathbf{y}\|_2 \leq \varepsilon \tag{2}$$

where $\mathbf{y}^T = [\mathbf{x}_1^T, \mathbf{x}_2^T, ..., \mathbf{x}_K^T]$, $\mathbf{x}_i$ is the representation of Y over $\mathbf{\Phi}_i$, and ε is a residual error. In theory, this optimization task is likely to lead to a successful separation of the mixed signal, such that $X_i \approx \mathbf{\Phi}_i \mathbf{x}_i$ [11].

The question is that whether such a method for signal separation can be used for speaker recognition. In speaker recognition task, the signal is not a mixture but the speech of one speaker, i.e., $Y = X_i$ with X_i denoting the speech of the ith speaker. If the ith ($i = 1, 2, ..., K$) speaker has a dictionary $\mathbf{\Phi}_i$, over which the speech of this speaker is sparse and the speech of other speakers is dense, we can construct $\mathbf{\Psi} = [\mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_K]$. In ideal conditions, solving equation (2) leads to $X_i \approx \mathbf{\Phi}_i \mathbf{x}_i$, i.e. only the atoms of $\mathbf{\Phi}_i$ are used. Therefore, we can classify Y via the label of the used atoms. In noisy environments, we can treat the noise as a speaker. Speaker and noise dictionaries which satisfy the assumption of MCA can be obtained by training with the corresponding signals. Therefore, it is feasible to apply MCA in the speaker recognition task.

3. Methodology

According to the above analysis, we can expect to achieve good performance by designing a method based on MCA. Before that, there are three problems needed to be considered:

1) How to train a big dictionary which satisfies the assumption of MCA. MCA can play as expected only when its assumption holds true in speaker recognition. The essential requirement of this assumption is to improve the discriminability of the big dictionary.

2) How to cope with varying noise. It is difficult to predict the noise in speech because it may have two kinds of change. The first is the change of noise type, e.g., change from one noise to another noise. The second is that a noise may be time-varying itself. A pre-trained noise dictionary without change will cause a mismatch to degrade the performance.

3) How to make classification with sparse representation. In ideal conditions, the speech of one speaker can only be sparsely represented over the dictionary of this speaker; however, in real applications, atoms of other speakers may also be used, resulting in errors in classification.

Focused on the above problems, we design a novel method for speaker recognition (shown in Fig. 1).

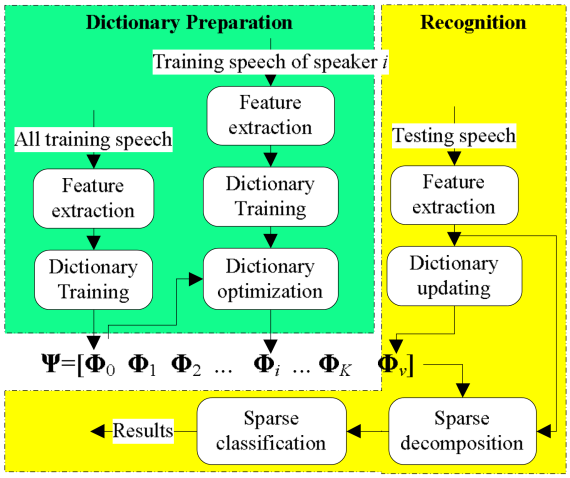

Figure 1: Procedure of speaker recognition based on MCA.

3.1. Dictionary preparation

3.1.1. Feature extraction

The used feature is the magnitude spectrum. In feature extraction, speech signals are first split into overlapping frames, and a Hamming window is added to each frame. Next, the discrete Fourier transform (DFT) is made on speech frames and the magnitude spectrum is computed as the input feature.

3.1.2. Structure of big dictionary

For a speaker recognition system with K enrolled speakers, we design a novel structure of the big dictionary:

$$\mathbf{\Psi} = [\mathbf{\Phi}_0, \mathbf{\Phi}_1, \mathbf{\Phi}_2, ..., \mathbf{\Phi}_K, \mathbf{\Phi}_v]. \tag{3}$$

$\mathbf{\Phi}_0$ is a universal background dictionary (UBD) which models the common variability shared by all speakers. Here we have borrowed the idea of the GMM-UBM, which employs a UBM to model background. $\mathbf{\Phi}_i (i = 1, ..., K)$ is the dictionary to model the variability of the ith speakers. $\mathbf{\Phi}_v$ is the noise dictionary to model environment noise. All atoms in $\mathbf{\Psi}$ are normalized to be unit-norm vectors.

Such a dictionary structure possesses two important advantages. The first is that it makes the big dictionary more discriminant. The second is that the atom number of each speaker dictionary can be reduced substantially, which reduces the computational load of sparse decomposition.

3.1.3. Dictionary training

Many methods are proposed to train a dictionary, e.g., the k-SVD [15], k-means [16]. In our method, the k-SVD is chosen to train dictionaries. The dictionary training problem is described as

$$\min \|\mathbf{Y} - \mathbf{\Phi}\mathbf{X}\|_2^2 \text{ subject to } \|\mathbf{x}_i\|_0 \leq T_0 \tag{4}$$

where $\mathbf{Y} = [Y_1, Y_2, ..., Y_M]$ is a training dataset with each Y_i being a feature vector of a speech frame. $\mathbf{\Phi}$ is the dictionary. $\mathbf{X} = [\mathbf{x}_1, \mathbf{x}_2, ..., \mathbf{x}_M]$ is the set of sparse vectors corresponding to $\mathbf{Y}$ and T_0 is the constraint on sparsity. UBD can be trained with large amounts of unlabelled speech from different speakers. Each $\mathbf{\Phi}_i$ is trained with the speech from the ith speaker, with $\mathbf{\Psi}$ as an initial.

3.1.4. Dictionary optimization

In the proposed method, we use $\mathbf{\Phi}_0$ to model the common variability shared by all speakers. We hope that the dictionary of each speaker only model the difference between this speaker and other speakers, which can improve discriminability of the dictionary. To this end, we provide a method to optimize speaker dictionaries (shown in Algorithm 1).

Algorithm 1 Dictionary optimization

Input: $\mathbf{\Phi}_0 = [\mathbf{a}_1, \mathbf{a}_2, ..., \mathbf{a}_{M_0}]$, $\mathbf{\Phi}_i = [\mathbf{b}_1, \mathbf{b}_2, ..., \mathbf{b}_{M_i}]$, T
Output: $\hat{\mathbf{\Phi}}_i$
1: **initialize**: $\mathbf{w} = [1, 1, ..., 1]$, $\hat{\mathbf{\Phi}}_i = [\,]$;
2: **for** p = 1 **to** M_i **do**
3: **for** q = 1 **to** M_0 **do**
4: **if** $(|\mathbf{b}_p^T \mathbf{a}_q| < \mathbf{w}[p])$ **then**
5: $\mathbf{w}[p] = |\mathbf{b}_p^T \mathbf{a}_q|$;
6: **end if**
7: **end for**
8: **end for**
9: γ = the Tth smallest element of $\mathbf{w}$;
10: **for** p = 1 **to** M_i **do**
11: **if** $\mathbf{w}[p] < \gamma$ **then**
12: $\hat{\mathbf{\Phi}}_i = [\hat{\mathbf{\Phi}}_i, \mathbf{b}_p]$;
13: **end if**
14:**end for**

When $|\mathbf{a}_i^T \mathbf{b}_j|$ approaches 1, $\mathbf{a}_i$ and $\mathbf{b}_j$ are prone to be the same; if approaches 0, they are completely different. The main idea of this algorithm is to remove atoms which are close to the atoms of the UBD.

3.2. Recognition

3.2.1. Sparse decomposition

Sparse decomposition is to solve the problem in the equation (2), which is proved to be NP-hard and is impossible to be solved by sweeping exhaustively through all possible sparse subsets. However, if $\mathbf{x}$ is sparse or approximately sparse, it can be uniquely determined by solving [17]

$$\mathbf{y} = \arg\min_{\mathbf{y}} \lambda \|\mathbf{y}\|_1 + \frac{1}{2} \|Y - \mathbf{\Psi}\mathbf{y}\|_2^2 \tag{5}$$

where $\lambda > 0$ is the regularization parameter. Equation (5) is also referred as BPDN, which is adopted in the proposed method.

3.2.2. Noise dictionary update

The unknown noise is time-varying; therefore, it is impossible to model it well using a fixed dictionary. A feasible method is to update the noise dictionary with the noisy speech for test. The update method is shown in algorithm 2, where $\mathbf{\Upsilon}$ is a test utterance corrupted by noise, δ is a threshold of sparse degree, $\mathbf{y_s}$ is the sparse representation of Y_i over $[\mathbf{\Phi}_0, ..., \mathbf{\Phi}_K]$, Γ is used to store noise observations, and $kSVD(\mathbf{\Phi}_v, \Gamma, M_v)$ denotes the function to train a dictionary with M_v atoms, using Γ as the training data and $\mathbf{\Phi}_v$ as an initial. The dictionary is trained in an on-line way. The training data are the residual of the noisy speech subtracting the clean speech. The training procedure is repeated until the sparse degree is converged.

In this algorithm, speech dictionary keeps unchanged. In the first iteration, all noisy speech frames are decomposed over the speech dictionary (noise dictionary is null). Since speech is sparser than noise, part of noise is in the residual error, with which the noise dictionary is trained. In the remaining iterations, the noise in the noisy speech is represented over the noise dictionary.

Algorithm 2 Noise dictionary update

Input: $\mathbf{\Psi} = [\mathbf{\Phi}_0, ..., \mathbf{\Phi}_K, \mathbf{\Phi}_v]$, where $\mathbf{\Phi}_v = [\mathbf{c}_1, \mathbf{c}_2, ..., \mathbf{c}_{M_v}]$, $\mathbf{\Upsilon} = [Y_1, Y_2, ..., Y_P]$, δ
Output: $\mathbf{\Phi}_v$
1: **initialize**: $\Gamma = [\,]$;$\varsigma = \delta$; $\mathbf{\Phi}_v = [\,]$;
2: **while** $\varsigma \geq \delta$ **do**
3: $\varsigma = 0$;
4: $\Gamma = [\,]$;
5: **for** i = 1 **to** P **do**
6: $\mathbf{y}$=argmin$_{\mathbf{y}}$ $\|\mathbf{y}\|_0$ subject to $\|Y_i - \mathbf{\Psi}\mathbf{y}\|_2 \leq \varepsilon$;
7: $\varsigma = \varsigma + \|\mathbf{y}\|_1$;
8: $\Gamma = [\Gamma,\ (Y_i - [\mathbf{\Phi}_0, ..., \mathbf{\Phi}_K]\mathbf{y}_s)]$;
9: **end for**
10: $\mathbf{\Phi}_v = kSVD(\mathbf{\Phi}_v, \Gamma, M_v)$;
11: $\varsigma = \varsigma / P$;
12:**end while**

3.2.3. Sparse classification

For a set of test speech frames $[Y_1, Y_2, ..., Y_P]$ with the corresponding sparse representations $[\mathbf{y}_1, \mathbf{y}_2, ..., \mathbf{y}_P]$, we assign this speech to one speaker according to

$$i = \arg \max_{1 \leq i \leq K} \frac{1}{P} \sum_{j=1}^{P} \|\delta_i(\mathbf{y}_j)\|_1 \tag{6}$$

where $\delta_i(.)$ gives us a vector $\in \mathbb{R}^N$ where the only nonzero entries are from the ith class. Note that only the representation coefficients over speaker dictionaries are counted. By this way, the influence of the common variability and the noise can be reduced.

4. Experiments and analysis

4.1. Experiment setup

The 863-database is used to evaluate the proposed method. This corpora is established from the Chinese National Hi-Tech Project for ASR system development. It contains 96269 sentences spoken by 83 males and 83 females (520 utterances for each speaker). The data are recorded in a noise-free environment through a close-talk microphone, with a sample rate of 16 kHz and 16-bit quantization. The first 10 utterances of each speaker are used for dictionary training, the next 10 utterances of each speaker are used for testing. Each test utterance is used as a trial and the recognition task is close-set. The data for training and testing are resampled at a rate of 8 kHz. The UBD which contains 512 atoms, is trained with all training data. The dictionary of each speaker is trained with the training data of this speaker, with the UBD as an initial, and then optimized via Algorithm 1. To obtain noisy test speech, four noises (taken from the Noisex-92 database [18]), namely white, f16, babble and pink, were added artificially on the test utterances at 0 dB, 5 dB, 10 dB and 20 dB. Then we have 4 noisy versions of the testset.

In the frontend, speech signals are Hamming windowed every 10 ms with a window width of 20 ms. The number of DFT points is 512 and the dimension of each observation is 257. The parameter T in Algorithm 1 is set as 64, which means that there are 64 atoms left for a speaker dictionary after optimization.

Table 1: Performance comparison under various noisy environments (recognition accuracy, %)

Methods	Clean	White (dB)				Pink (dB)				F16 (dB)				Babble (dB)			
		0	5	10	20	0	5	10	20	0	5	10	20	0	5	10	20
GU	96.4	6.0	20.5	51.2	78.3	9.0	25.9	49.4	80.1	13.2	31.9	48.8	85.5	21.7	31.9	54.2	87.3
GS	97.8	13.6	37.7	62.8	80.5	19.2	33.4	53.6	83.5	16.6	34.8	51.3	85.0	22.7	36.1	54.8	88.7
PWF	96.2	60.5	70.5	83.0	93.8	53.6	70.3	84.2	91.4	62.0	69.5	85.6	88.3	63.5	67.3	80.2	91.6
NNE	96.2	61.3	68.3	83.2	94.2	58.7	71.5	85.8	93.1	62.2	70.4	86.2	89.5	63.2	67.8	80.2	90.9
SparseA	**98.2**	**62.9**	**70.5**	**83.4**	**94.5**	**61.6**	**74.0**	**85.0**	**90.8**	**60.6**	**69.9**	**85.7**	**90.1**	**63.9**	**65.2**	**80.3**	**92.2**
SparseB	**97.6**	**65.4**	**74.3**	**84.3**	**94.6**	**70.2**	**77.1**	**86.7**	**93.4**	**62.4**	**73.2**	**88.0**	**90.3**	**65.2**	**69.9**	**80.7**	**92.2**

The atom number of the noise dictionary is also 64. The λ in (5) is set as 0.01 and the δ in algorithm 2 is set as 7. The baselines for comparison are the GMM-UBM and GMM-SVM without compensation. The used features are 13-dimensional MFCC coefficients ($c_0 \sim c_{12}$), appended by their first- and second-order derivatives. The cepstral mean and variance normalization are applied to the feature. The UBM consists of 1024 mixture components, which are trained using the expectation-maximization algorithm with all the training data. The GMM of each speaker is trained with the training data of this speaker via maximum-a-posteriori adaptation. As for GMM-SVM, each training utterance is transformed a GMM mean supervector, and all supervectors are used to train a SVM classifier. Another methods chosen for comparison is the perceptual wiener lter (PWF) [19] and non-stationary noise estimation (NNE) [20], which enhance speech and then extract feature from enhanced speech for recognition with GMM-UBM. The measure of the performance is the recognition accuracy.

4.2. Results and analysis

The results are shown in Table 1, where GU denotes the GMM-UBM, GS denotes the GMM-SVM, PWF denotes the perceptual Wiener filtering, NNE denotes the non-stationary noise estimation, SparseA denotes the proposed method without using noise dictionary, and SparseB denotes the proposed method using noise dictionary. We can see that all methods can achieve satisfied performances under clean conditions. The performance of the GMM-UBM decreases rapidly with the increase of noise level. Its accuracy falls down to 6.0 % at 0 dB of white noise. The performance of GMM-SVM is better than that of the GMM-UBM, but it is also sensitive to noises. The improvements of the PWF and NNE are obvious. SparseA also shows robustness to noise although it does not use a noise dictionary. SparseB has achieved better performance than other methods. When the SNR is 20 dB, the accuracy can reach upto more than 90%; when the SNR is low to 0 dB, the accuracy can keep upto 60 %.

5. Conclusion

The performance of speaker recognition systems decreases rapidly in noisy environments; therefore, the noise robustness plays an important role. Current research mainly focuses on channel distortion and session variability. There is lack of valid method to overcome the influence of noise. To solve this problem, we propose a speaker recognition method based on MCA. We analyze the assumption of MCA and attempt to design a method to make this assumption true. First, we train dictionaries in the magnitude spectrum domain, where speech is prone to be sparse. We prepare dictionaries via the training method rather than collecting exemplars. Second, we design a novel dictionary structure and optimize dictionaries in a discriminant way. On the one hand, this method allows the speaker dictionaries only modelling the difference between speakers; on the other hand, the atom number can be reduced substantially. Third, we use a noise dictionary in the big dictionary and provide an algorithm to update this dictionary according to noisy conditions, which ensures varied noise can be represented and absorbed. Finally, we simulate noisy environments by artificially adding noise on clean speech to validate the proposed method. The results show that the proposed method is robust to various noises.

6. References

[1] D. Reynolds, T.F. Quatieri, and R.B. Dunn, "Speaker verification using adapted Gaussian mixture models," *Digital Signal Process.*, vol. 10, pp. 19–41, 2000.

[2] W.M. Campbell, D.E. Sturim, and D.A. Reynolds, "Support vector machines using GMM supervectors for speaker verification, " *IEEE Signal Processing Letters*, vol. 13, pp. 308–311, 2006.

[3] P. Kenny, G. Boulianne, P. Ouellet, and P. Dumouchel, "Joint factor analysis versus eigenchannels in speaker recognition," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1435–1447, May 2007.

[4] J. Ming, T.J. Hazen, J.R. Glass, and D.A. Reynolds, "Robust speaker recognition in noisy conditions," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 5, pp. 1711–1723, May 2007.

[5] B. Atal, "Effectiveness of linear prediction characteristics of the speech wave for automatic speaker identification and verification," *J. Acoust. Soc. Am.,* vol. 55, no. 6, 1304–1312. 1974.

[6] O. Viikki, D. Bye, and K. Laurila, "A recursive feature vector normalization approach for robust speech recognition in noise," *in Proc. ICASSP*, 1998, pp. 733–736.

[7] A. de la Torre, A. M. Peinado, J. C.Segura, J.L.Perez-Cordoba, M. C. Benitez, and A. J. Rubio, "Histogram equalization of speech representation for robust speech recognition," *IEEE Trans. Speech Audio Process.*, vol. 13, no. 3, pp. 355-366, 2005.

[8] C. P. Chen, J. A. Bilmes, "MVA Processing of Speech Features," *IEEE Transactions on Audio, Speech, and Language Processing*, vol. 15, no. 1, pp. 257–270, Jan. 2007.

[9] T. F. Quatieri, Discrete-Time Speech Signal Processing: Principles and Practice. Upper Saddle River, NJ, USA: Prentice-Hall, 2001.

[10] D.L. Donoho "Compressed sensing," *IEEE Trans. on Inf. Theory*, vol. 52, no. 4, pp. 12891306, Apr. 2006.

[11] J. Bobin, J.-L. Starck, J.M. Fadili, Y. Moudden, and D.L. Donoho, "Morphological component analysis: An adaptive thresholding strategy," *IEEE Trans. on Image Processing*, vol. 16, no. 11, pp. 2675–2681, 2007.

[12] I. Naseem, R. Togneri, and M. Bennamoun, "Sparse representation for speaker identification," in Proc. *ICPR*, 2010, pp. 4460–4463.

[13] J.M.K. Kua, E. Ambikairajah, J. Epps,and R. Togneri, "Speaker verification using sparse representation classification," in Proc. *ICASSP*, 2011, pp. 45484551.

[14] N. Dehak, P. Kenny, R. Dehak, P. Dumouchel, and P. Ouellet, "Front-end factor analysis for speaker verification," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 19, no.4, pp. 788–798, May 2010.

[15] M. Aharon, M. Elad, and A.M. Bruckstein, "The K-SVD: An algorithm for designing of overcomplete dictionaries for sparse representation, " *IEEE Trans. on Signal Process.*, vol. 54, no. 11, pp. 4311–4322, Nov. 2006.

[16] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," In Proc. *BSMSP*, Berkeley, University of California Press, 1967, pp. 281–297.

[17] S. Chen, D. Donoho and M. Saunders, "Atomic decomposition by basis pursuit," *SIAM Rev.*, vol. 43, no. 1, pp.129–159, 2001.

[18] [Online]. Available: http://www.speech.cs.cmu.edu/comp.speech/Section1/Data/noisex.html.

[19] L. Lin, W. Holmes, and E. Ambikairajah, "Subband noise estimation for speech enhancement using a perceptual wiener lter," in *Proc. ICASSP*, 2003, vol. 1, pp. 8083.

[20] S. Rangachari and P. Loizou, "A noise-estimation algorithm for highly non-stationary environments," *Speech Commun.*, vol. 48, pp. 220231, 2006.