An Improved Dictionary Learning Method for Speech Enhancement

Yue Hao, Changchun Bao
Speech and Audio Signal Processing Laboratory, School of Electronic Information and Control Engineering, Beijing University of Technology, Beijing, China, 100124
E-mail: S201302071@emails.bjut.edu.cn, baochch@bjut.edu.cn

Abstract— **In this paper, an improved dictionary learning method for speech enhancement is proposed. Given prior information of the noise, the dictionaries of speech and noise are firstly trained by an approximate KSVD algorithm, respectively. Then, the estimated short-time Fourier transform (STFT) magnitudes of speech and noise can be sparsely represented by multiplying the dictionary with sparse coefficients, which are calculated by the least angle regression (LAR) algorithm. A geometrical stopping criterion with an adaptive threshold is utilized to adjust the conventional stopping criterion in LAR algorithm so that it can increase the adaptability of LAR. Next, we propose a framework that utilizes the expectation maximization (EM) method to refine the energy of the estimated speech and noise in order to obtain more accurate estimation of STFT magnitudes. Finally, a modified wiener filter is constructed to further eliminate residual noise. When the prior information of noise is unknown, an online noise estimation method is applied to replace the noise dictionary. The test results show that the proposed method outperforms the reference speech enhancement methods.**

Index Terms—Speech enhancement, Dictionary learning, Sparse representation, EM framework, Modified Wiener filtering, Noise estimation

I. Introduction

The goal of speech enhancement is to remove noise from noisy speech for improving speech quality and intelligibility. Conventional single channel speech enhancement methods, such as Wiener filtering [1], spectral-subtraction method [2], and short-time spectral amplitude (STSA) estimators [3], do not explicitly model the characteristics of speech as a priori information, which leads to a poor performance of speech enhancement under non-stationary noise environment. To solve this problem, speech-specific information is incorporated as a priori information, which has been applied in a wide range of methods, including codebook-based (CB) methods [4-5] and dictionary learning (DL) methods [6]. These methods can achieve better performance for noise suppression and speech enhancement. For CB method, the trained linear predictive coefficients (LPC) codebooks of both speech and noise are used as a priori information. In [4-5], either one pair [4] or all pairs [5] of code-word vectors of speech and noise are selected, and the corresponding gains are estimated online based on the minimum Itakura-Saito distortion criterion. The main shortage of these methods lies in two aspects. On one hand, there are too many sparse representations so it may be easy to induce an approximation error, which can be found in maximum-likelihood method [4]. On the other hand, although a high dense representation avoids the approximation error, it results in source ambiguity [7], which refers to Bayesian method [5].

For DL method [6], the dictionaries of speech and noise using approximate KSVD algorithm [8] are trained as a priori information, which is an extension of CB method. The short-time Fourier transform (STFT) magnitudes of speech and noise are approximated by sparse linear combination of multiple atoms based on the minimum approximation error criterion. The advantage of this method is that it achieves a tradeoff between approximation error and source ambiguity. However, there are also some problems. (1) The sparse representation of speech is often not accurate using a pre-set threshold, which is adopted in the conventional stopping criterion (CSC) for least angle regression (LAR) algorithm. (2) Each of multiple noise dictionaries is trained for one type of noise, which doesn't meet the demands of practical application as well as increases the complexity.

In order to address the aforementioned problems, an improved dictionary learning method is proposed. The main contribution of this paper is described as follows. Firstly, we achieve a more accurate sparse representation of speech by using a geometrical stopping criterion (GSC) with an adaptive threshold [9] instead of a fixed value in the LAR algorithm. Secondly, a framework with expectation maximization (EM) algorithm [10] is proposed to refine the energy of estimated speech and noise for reducing further distortions. Thirdly, the estimated speech and noise are used to construct a modified wiener filter (MWF) that exploits normalized cross-correlation coefficients (NCCC) [11] between the spectra of noisy speech and noise, so the fluctuant background noise can be depressed as much as possible. Lastly, we use the minima controlled recursive averaging (MCRA) [12] algorithm for online noise estimation to replace the noise dictionary, so that we do not rely on the noise classification any more. Experimental results demonstrate that the proposed method outperforms the reference methods.

The remainder of this paper is organized as follows. Section II presents an overview of conventional dictionary learning (CDL) method proposed in [6]. The proposed improved DL method is described in Section III. The performance evaluation results are shown in Section IV and Section V gives the conclusions.

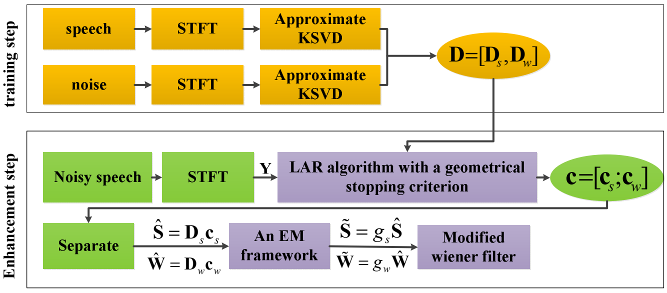
training step
speech
STFT
Approximate KSVD
noise
STFT
Approximate KSVD
$\mathbf{D}=[\mathbf{D}_s,\mathbf{D}_w]$
Enhancement step
Noisy speech
STFT
$\mathbf{Y}$
LAR algorithm with a geometrical stopping criterion
$\mathbf{c}=[\mathbf{c}_s;\mathbf{c}_w]$
Separate
$\hat{\mathbf{S}}=\mathbf{D}_s\mathbf{c}_s$
$\hat{\mathbf{W}}=\mathbf{D}_w\mathbf{c}_w$
An EM framework
$\tilde{\mathbf{S}}=g_s\hat{\mathbf{S}}$
$\tilde{\mathbf{W}}=g_w\hat{\mathbf{W}}$
Modified wiener filter

Fig.1. The block diagram of the proposed speech enhancement method.

II. CONVENTIONAL DICTIONARY LEARNING

Let $\mathbf{Y}\in\mathbb{R}^K$, $\mathbf{S}\in\mathbb{R}^K$ and $\mathbf{W}\in\mathbb{R}^K$ denote STFT magnitudes of noisy speech, clean speech and noise, respectively. Considering an additive noise model in STFT magnitude domain,where speech and noise are independent, **Y**=**S**+**W**, we define the generative dictionary as: $\mathbf{D}=[\mathbf{D}_s,\mathbf{D}_w]\in\mathbb{R}^{K*(L_s+L_w)}$, where $\mathbf{D}_s$ is the normalized speech dictionary with L_s atoms and $\mathbf{D}_w$ is the normalized noise dictionary with L_w atoms.

During the enhancement step, the goal of DL is to estimate the sparse coefficients of speech $\mathbf{c}_s\in\mathbb{R}^{L_s}$ and noise $\mathbf{c}_w\in\mathbb{R}^{L_w}$. These coefficients can be obtained by minimizing the approximation error [6]

$$\mathbf{c}=\arg\min_{\mathbf{c}}\left\|\mathbf{Y}-\mathbf{D}\mathbf{c}\right\|_F^2 \tag{1}$$

where $\mathbf{c}=[\mathbf{c}_s;\mathbf{c}_w]$, and **D** is trained offline based on the approximate KSVD algorithm [8]. The LAR algorithm is adopted to give an approximate solution to (1), by using CSC with a pre-set threshold u_{thresh} [6]

$$\frac{\max|\mathbf{D}^T\mathbf{e}^{(i)}|}{\left\|\mathbf{e}^{(i)}\right\|}<u_{thresh} \tag{2}$$

where $\mathbf{e}^{(i)}$ represents the approximation error at step i in the iterative process, $\mathbf{e}^{(i)}=\mathbf{Y}^{(i)}-\hat{\mathbf{Y}}^{(i)}$, $\hat{\mathbf{Y}}^{(i)}$ is the estimated STFT magnitude vector.

Once **c** is obtained, the estimated STFT magnitudes of speech and noise can be expressed as $\hat{\mathbf{S}}=\mathbf{D}_s\mathbf{c}_s$ and $\hat{\mathbf{W}}=\mathbf{D}_w\mathbf{c}_w$, respectively. And $\hat{\mathbf{S}}$ and $\hat{\mathbf{W}}$ can be used to construct a geometric spectral subtraction filter [6] to obtain the enhanced speech.

III. IMPROVED DICTIONARY LEARNING

In this section, we propose an improved dictionary learning method. Given prior knowledge of the noise, the overall block diagram of the proposed speech enhancement method is illustrated in Fig.1. In the training step, the time domain signals of speech and noise are transformed into the STFT magnitude domain. And dictionaries of both speech and noise are trained offline based on the approximate KSVD algorithm. Both $\mathbf{D}_s$ and $\mathbf{D}_w$ are combined into **D** for speech enhancement. In the enhancement step, the noisy speech is firstly transformed into the STFT magnitude domain. Then, **Y** is sparsely represented in **D** using LAR algorithm with GSC (Section III-A). The sparse coefficient **c** is separated into $\mathbf{c}_s$ and $\mathbf{c}_w$, in order to obtain $\hat{\mathbf{S}}$ and $\hat{\mathbf{W}}$. Next, a framework with EM algorithm is applied to refine the energy of $\hat{\mathbf{S}}$ and $\hat{\mathbf{W}}$ (Section III-B). Lastly, a MWF is constructed based on the refined $\tilde{\mathbf{S}}$ and $\tilde{\mathbf{W}}$ (Section III-C). We can obtain the magnitude spectrum of the enhanced speech by filtering the noisy speech through the MWF. Moreover, the case without the prior knowledge of noise is described in Section III-D.

A. A geometrical stopping criterion for LAR algorithm

The CDL method utilizes LAR algorithm with a pre-set threshold in (2), which influences the accuracy of sparse representation for different signals. Therefore, CDL method has insufficient adaptability. The choice of threshold according to GSC [9] does not require pre-set threshold. It was firstly proposed with a Volterra filter for nonlinear system identification, which has been reported closer to real scenarios. In this section, the GSC with an adaptive threshold is adopted to obtain **c**. It can change dynamically according to the speech signals with different noise levels and types.

We define the correlation vector $\boldsymbol{\beta}\in\mathbb{R}^{L_s+L_w}$ between dictionary matrix and approximation error as $\boldsymbol{\beta}=\mathbf{D}^T\mathbf{e}^{(i)}$. The j^{th} element at step i can be written as $\beta_j^{(i)}=\mathbf{d}_j^T\mathbf{e}^{(i)}=\|\mathbf{d}_j^T\|\|\mathbf{e}^{(i)}\|\cos\theta_j^{(i)}$. Since the j^{th} column-vector $\mathbf{d}_j$ of matrix **D** is normalized during the training step, we have

$$\theta_j^{(i)}=\arccos\frac{\beta_j^{(i)}}{\left\|\mathbf{e}^{(i)}\right\|} \tag{3}$$

The LAR algorithm is stopped when $\Delta\boldsymbol{\theta}^{(i)}$ reaches an adaptive threshold. That is, $\Delta\boldsymbol{\theta}^{(i)}\le\sigma_{\boldsymbol{\theta}^{(1)}}$, where $\Delta\boldsymbol{\theta}^{(i)}=\max(\boldsymbol{\theta}^{(i)})-\min(\boldsymbol{\theta}^{(i)})$, $\boldsymbol{\theta}^{(i)}=[\theta_1^{(i)},\theta_2^{(i)},...,\theta_j^{(i)},...,\theta_{L_s+L_w}^{(i)}]$ is the angle vector at step i, $\sigma_{\boldsymbol{\theta}^{(1)}}$ is the standard deviation of the angles at the first step. The new stopping criterion is used to replace CSC represented by (2) to obtain a more accurate solution to (1). And this new algorithm with an adaptive threshold can more effectively track the energy changes of the target speech and noise.

B. An EM framework for refining dictionary learning

In terms of the CDL method, the target **S** and **W** cannot be separately sparse-represented well, especially under speech-like non-stationary noise environment. The main reason is that there is always a risk that parts of the estimated speech component are tend to be represented by speech-like noise atoms [6]. A framework using EM algorithm is proposed to refine the energy of the estimated $\hat{\mathbf{S}}$ and $\hat{\mathbf{W}}$ so that the estimation of STFT magnitudes will be more accurate. According to the theory that the energy compensation factor, $\mathbf{g}=\{g_s,g_w\}$, can be considered as the parameter for estimation, the refined STFT magnitudes of speech and noise can be expressed as $\tilde{\mathbf{S}}=g_s\hat{\mathbf{S}}$ and $\tilde{\mathbf{W}}=g_w\hat{\mathbf{W}}$, respectively.

Therefore, the m^{th} iteration of the EM algorithm can be described in the following two steps.

(a)E-step involves the evaluation of the auxiliary function

$$Q(\mathbf{g}\,|\,\mathbf{g}^{(m-1)})=\mathrm{E}\{\log p(\mathbf{S},\mathbf{W}\,|\,\mathbf{g})\,|\,\mathbf{Y},\mathbf{g}^{(m-1)}\} \tag{4}$$

where $\mathbf{g}^{(m\text{-}1)}$ denotes the estimated parameter for $(m\text{-}1)^{\text{th}}$

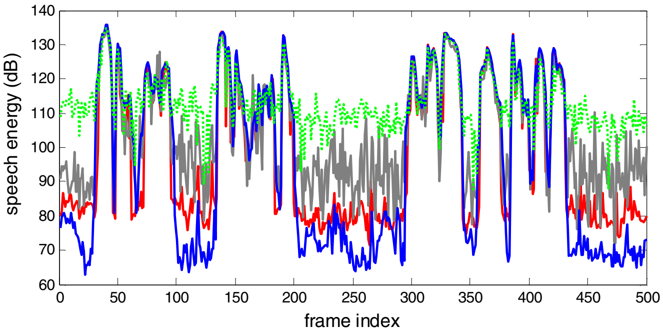

Fig.2. The comparison of STFT energy for four cases. The blue line is the STFT energy of original clean speech. The green line is the STFT energy derived from CDL method. The gray line denotes the STFT energy derived from LAR algorithm with GSC, and the red line is STFT energy derived from the proposed framework with EM algorithm.

iteration. Since **S** and **W** are independent, Eq. (4) can be simplified as

$$Q(\mathbf{g}\mid\mathbf{g}^{(m-1)})\approx \int p(\mathbf{S}\mid\mathbf{Y},g_s^{(m-1)})\log p(\mathbf{S}\mid g_s)\,\mathrm{d}\mathbf{S}+\int p(\mathbf{W}\mid\mathbf{Y},g_w^{(m-1)})\log p(\mathbf{W}\mid g_w)\,\mathrm{d}\mathbf{W} \tag{5}$$

Under the Gaussian assumptions of Re{**S**}, Im{**S**}, Re{**W**}, and Im{**W**}, the probability density function (PDF) of **S** and **W** given **g** is Rayleigh [13]

$$p(\mathbf{S}\mid g_s)=\frac{2^K\,|\operatorname{diag}\mathbf{S}|}{g_s\,|\mathbf{V}_s|}\exp[-\mathbf{S}^T(g_s\,\mathbf{V}_s)^{-1}\mathbf{S}] \tag{6}$$

where $\mathbf{V}_s$ is a diagonal covariance matrix of speech, and the k^{th} diagonal elements of $\mathbf{V}_s$ is calculated by $\mathrm{E}\{|\hat{\mathbf{S}}(\omega_k)|^2\}$, where k denotes the frequency bin. The PDF of **W** can be calculated similarly as (6).

(b)M-step determines **g** that maximizes $Q(\mathbf{g}|\mathbf{g}^{(m\text{-}1)})$

$$\begin{aligned} g_s^{(m)}&=\int p(\mathbf{S}\mid\mathbf{Y},g_s^{(m-1)})(\mathbf{S}^T\mathbf{V}_s^{-1}\mathbf{S})\,\mathrm{d}\mathbf{S}\\ g_w^{(m)}&=\int p(\mathbf{W}\mid\mathbf{Y},g_w^{(m-1)})(\mathbf{W}^T\mathbf{V}_w^{-1}\mathbf{W})\,\mathrm{d}\mathbf{W}\end{aligned} \tag{7}$$

Fig.2 gives an example for comparing the STFT energy corrupted by babble noise with 5dB. It can be seen that the proposed LAR algorithm with GSC can track the target speech more effectively than CDL method, with a little distortion for a few frames. And, the proposed framework with EM algorithm can not only refine these distortions, but also track the target speech more quickly than LAR algorithm with GSC in sudden change regions of speech energy. As a result, we can effectively remove more noise and retain more speech components at the same time.

C. *Modified wiener filter for eliminating residual noise*

To further eliminate residual noise in unvoiced or silence segments, the NCCC denoted by ρ is exploited in an enhancement filter, where the STFT magnitudes of speech and noise are estimated by the framework given in Section III-B. We consider a conventional wiener filter (CWF) as $H(\omega)=|S(\omega)|^2/(|S(\omega)|^2+|W(\omega)|^2)$. We utilize ρ to modify the transfer function of CWF

$$H(\omega)=\frac{(1-\rho)\,|\tilde{S}(\omega)|^2}{(1-\rho)\,|\tilde{S}(\omega)|^2+\rho\,|\tilde{W}(\omega)|^2} \tag{8}$$

where ρ denotes the NCCC between the spectra of the noisy speech and estimated noise, which is given in [11].

D. *Dictionary learning without prior information of noise*

In the real environment, since the types of noise cannot be known in advance, the method relying on the noise classification is not suitable for the practical application. In order to fix this issue, a robust noise estimation method is embedded in the proposed DL method. In this way, only speech dictionary need to be trained.

During the enhancement step of DL, we firstly obtain $\mathbf{c}_s$ by solving the following optimization problem

$$\mathbf{c}_s=\arg\min_{\mathbf{c}_s}\left\|\mathbf{Y}-\mathbf{D}_s\mathbf{c}_s-\sqrt{\sigma_w^2}\right\|_F^2 \tag{9}$$

where σ_w^2 is the estimated noise power spectrum, which is derived from MCRA algorithm [12] instead of multiplying the noise dictionary with sparse coefficients of the noise. An approximated solution to (9) can be obtained based on the proposed LAR algorithm with GSC, which is similar to Section III-A. Secondly, the framework with EM algorithm given in Section III-B is applied to get a better estimation, due to the inaccurate estimation of MCRA algorithm under non-stationary noise environment. Lastly, the MWF in Section III-C is used to further eliminate residual noise.

IV. PERFORMANCE EVALUATION

We evaluated the performance of the proposed improved DL method, including two cases with and without prior information of the noise. For reference simplicity, PM1 and PM2 denote the first and second case, respectively. For both cases, the speech dictionary is trained with one hour speech database, i.e. $\mathbf{D}_s$, with L_s=1024 atoms. The number of frequency bins per frame in STFT magnitude domain is K=256. The frame length is 20ms with a frame shift of 10ms. The test speech is chosen from NTT database including 9 sentences from 4 female speakers and 5 male speakers. The length of each sentence sampled at 8kHz is 8s. In our experiments, noise signals are selected from Noisex-92 including white noise, babble noise, street noise and office noise. For PM1 case, the size of each noise dictionary is set to L_w=1024 atoms. The input signal to noise ratio (SNR) is defined as 0dB, 5dB, and 10dB, respectively.

The proposed method (PM) of two cases is compared with two reference methods, including CB method [4] and CDL method [6]. We utilize three objective evaluation measures to evaluate these methods, i.e. the average segmental signal-to-noise ratio (SSNR) [14], average log-spectral distortion (LSD) [15], and perceptual evaluation of speech quality (PESQ) measures [16]. In CB method [4], the speech and noise codebook sizes are 1024 and 8 respectively, except for babble noise with codebook size equal to 16. Table I, II, and III show the results of PESQ, LSD and SSNR, respectively.

From the Tables, it is readily seen that PM1 dramatically outperforms CDL and CB methods in all three test results, across all considered cases. This superior performance emphasizes the advantage of the proposed method on better sparse representation and more background noise removal.

Table I Results of PESQ

Noise type	Input SNR	Method				
		Noisy	CB	CDL	PM1	PM2
babble	0dB	1.80	1.72	2.16	2.27	2.05
	5dB	2.13	2.13	2.46	2.58	2.46
	10dB	2.50	2.47	2.75	2.81	2.83
street	0dB	2.30	2.59	2.79	2.96	2.82
	5dB	2.65	2.86	3.03	3.20	3.08
	10dB	2.95	3.09	3.28	3.45	3.3
office	0dB	2.02	2.16	2.56	2.69	2.53
	5dB	2.41	2.54	2.81	2.96	2.80
	10dB	2.76	2.84	3.07	3.19	3.13
white	0dB	1.36	1.72	1.64	1.86	2.11
	5dB	1.59	2.14	2.02	2.16	2.50
	10dB	1.97	2.40	2.35	2.42	2.81

Table II Results of LSD

Noise type	Input SNR	Method				
		Noisy	CB	CDL	PM1	PM2
babble	0dB	14.89	11.92	8.85	7.06	9.31
	5dB	12.84	10.30	7.57	5.66	7.48
	10dB	10.93	8.82	6.23	4.43	5.27
street	0dB	12.85	9.73	6.44	5.09	5.95
	5dB	10.93	8.31	5.21	3.99	4.53
	10dB	9.18	7.00	4.10	3.08	3.20
office	0dB	13.31	10.74	7.33	5.39	7.59
	5dB	11.35	9.20	5.98	4.19	6.05
	10dB	9.54	7.79	4.75	3.28	4.18
white	0dB	19.45	13.78	15.04	9.88	9.08
	5dB	17.19	12.03	13.05	8.69	7.98
	10dB	14.75	10.45	11.02	7.36	6.71

Table III Results of SSNR Improvements

Noise type	Input SNR	Method				
		Noisy	CB	CDL	PM1	PM2
babble	0dB	-	5.37	12.60	17.13	12.00
	5dB	-	4.81	10.88	15.61	9.92
	10dB	-	4.13	9.36	13.93	8.86
street	0dB	-	11.06	18.22	20.69	18.59
	5dB	-	9.95	16.43	18.91	17.75
	10dB	-	8.77	14.30	16.80	15.54
office	0dB	-	7.46	16.18	19.25	16.03
	5dB	-	6.65	15.26	18.24	14.94
	10dB	-	5.78	13.48	16.32	13.67
white	0dB	-	7.64	7.66	15.80	19.60
	5dB	-	7.18	6.65	14.56	17.76
	10dB	-	6.57	5.39	12.94	15.30

For PM2, it is better than reference methods with street and white noise for each scenario. There're some exceptions for office and babble noise. The main reason is that the estimated noise using MCRA algorithm is more accurate for white noise but slightly inaccurate under low SNR regions of non-stationary noise. However, since the difference is negligible, PM2 is much more preferred in the real world scenario, for not using noise dictionary.

V. CONCLUSIONS

In this paper, an improved DL method for speech enhancement is proposed. Given the prior knowledge of the noise, the PM has advantages of better sparse representation and more accurate energy level of speech and noise, which translates into significantly higher quality than that of reference methods. When the prior knowledge of noise is unknown, the PM can achieve slightly better results than reference methods without using any noise dictionary. This improves the robustness of real-world noise adaption.

ACKNOWLEDGMENT

This work was supported by the National Natural Science Foundation of China (Grant No.61471014).

REFERENCES

[1] J. S. Lim and A. V. Oppenheim, “Enhancement and bandwidth compression of noisy speech,” *Proc. IEEE,* vol. 67, no. 12, pp. 1586-1604, Dec. 1979.

[2] S. F. Boll, “Suppression of acoustic noise in speech using spectral subtraction,” *IEEE Trans. Acoust. Speech, Signal Processing,* vol. 27, no. 2, pp. 113-120, 1979.

[3] Y. Ephraim and D. Malah, “Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator,” *IEEE Trans. Acoust. Speech, Signal Processing,* vol. 32, no. 6, pp. 1109-1121, 1984.

[4] S. Srinivasan, J. Samuelsson, and W. B. Kleijn, “Codebook driven short-term predictor parameter estimation for speech enhancement,” *IEEE Trans. Audio, Speech, Lang. Processing*, vol. 14, no. 1, pp. 163–176, Jan. 2006.

[5] S. Srinivasan, J. Samuelsson, and W. B. Klejin, “Codebook-based bayesian speech enhancement for nonstationary environments,” *IEEE Trans. Acoust. Speech, Signal Processing,* vol. 15, no. 2, pp. 441-451, 2007.

[6] C. D. Sigg, T. Dikk, and J. Buhmann, “Speech enhancement using generative dictionary learning,” *IEEE Trans. Acoust. Speech, Signal Processing*, vol. 20, pp. 1698-1712, 2012.

[7] C. D. Sigg, T. Dikk, and J. Buhmann, “Speech enhancement with sparse coding in learned dictionaries,” in *Proc. IEEE Int. Conf. Acoust. Speech, Signal Process*, pp. 4758-4761, 2010.

[8] R. Rubinsterin, M. Zibulevsky, and M. Elad, “Efficient implementation of the K-SVD algorithm using batch orthogonal matching pursuit. Technical report,” Technion, Haifa, 2008.

[9] C. Valdman, M. L. R. de Campos, and J. A. Apolina´rio Jr, “A geometrical stopping criterion for the LAR algorithm,” i*n 20th European Signal Processing Conference(EUSIPCO)*, pp. 2104–2108, Aug. 2012.

[10] J. A. Bilmes, “A gentle tutorial of the EM algorithm and its application to parameter estimation for Gaussian mixture and hidden Markov models,” Univ. of Berkeley, Berkeley, CA, Tech. Rep. ICSI-TR-97-021, 1997.

[11] F. Bao, D. H. Jing, M. S. Jia, and C. C. Bao, “Speech enhancement based on a few shapes of speech spectrum,” *IEEE CHINASIP,* pp. 90-94, 2014.

[12] I. Cohen and B. Berdugo, “Noise estimation by minima controlled recursive averaging for robust speech enhancement,” *Signal Processing Letters, IEEE,* vol. 9, no. 1, pp. 12–15, 2002.

[13] A. Papoulis and S. Pillai, *Probablity, Random Variables and Stochastic Processed*. 4th ed. New York: McGraw Hill, Inc, 2002.

[14] S. R. Quackenbush, T. P. Barnwell, M. A. Clements, *Objective Measures of Speech Quality*. Englewood Cliffs, NJ: Prentice Hall, 1988.

[15] A. Abramson and I. Cohen, “Simultaneous detection and estimation approach for speech enhancement,” *IEEE Trans. Acoust. Speech, Signal Processing,* vol. 15, no. 8, pp. 2348-2359, 2007.

[16] ITU-T, Recommendation P.862, “Perceptual evaluation of speech quality (PESQ)”, 2001.