

## Speech enhancement based on hidden Markov model using sparse code shrinkage

E. Golrasan\* and H. Sameti

Department of Computer Engineering, Sharif University of Technology, Tehran, Iran.

Received 19 August 2015; Accepted 31 May 2016

\*Corresponding author: egolrasan@yahoo.com (E. Golrasan).

### Abstract

This paper presents a new hidden Markov model-based (HMM-based) speech enhancement framework based on the independent component analysis (ICA). We propose analytical procedures for training clean speech and noise models using the Baum re-estimation algorithm, and present a maximum a posteriori (MAP) estimator based on the Laplace-Gaussian (for clean speech and noise, respectively) combination in the HMM framework, namely sparse code shrinkage-HMM (SCS-HMM).

The proposed method on the TIMIT database in the presence of three noise types at three SNR levels in terms of PESQ and SNR are evaluated and compared with Auto-Regressive HMM (AR-HMM) and speech enhancement based on HMM with discrete cosine transform (DCT) coefficients using the Laplace and Gaussian distributions (LaGa-HMM<sub>DCT</sub>). The results obtained confirm the superiority of the SCS-HMM method in the presence of non-stationary noises compared to LaGa-HMM<sub>DCT</sub>. The results of the SCS-HMM method represent a better performance of this method compared to AR-HMM in the presence of white noise based on the PESQ measure.

**Keywords:** *Speech Enhancement, HMM-based Speech Enhancement, Multivariate Laplace Distribution, Independent Component Analysis (ICA transform), Sparse Code Shrinkage Enhancement Method.*

### 1. Introduction

Speech enhancement aims to improve speech quality using various algorithms. Enhancing speech degraded by noise, or noise reduction, is the most important field of speech enhancement, and is used for many applications such as mobile phones, VoIP, teleconferencing systems, speech recognition, and hearing aids.

Among the different proposed solutions, the statistical approach in speech enhancement is often preferred due to the stochastic nature of speech signals [1]. Generally, the statistical methods are divided into the model-based [2, 3] and non-model-based [4, 5] techniques. In the model-based procedures, the clean speech and noise models are first generated in a training phase, and then the clean speech is estimated based on this prior information in a test phase. The non-model-based procedures only consist of the test phase, and the required information is estimated using the noisy speech. Under the non-stationary noisy conditions, the model-based

techniques have advantages over the non-model-based techniques through prior information [3].

Hidden Markov Model (HMM) is one of the powerful model-based methods applied to speech enhancement and has resulted in high efficiency, especially under non-stationary noisy conditions [2]. One of the most important factors that influences the model precision of an HMM is the probability density function (pdf) of clean speech, noise, and noisy speech. In the HMM-based speech enhancement, the Gaussian pdf is used to model clean speech and noise, while the recent studies [4,6,7] have shown that clean speech and noise pdf are non-Gaussian distributions. The multivariate Laplace distribution has been recommended for modeling HMM as a non-Gaussian distribution [8]. In this modeling, using multivariate Laplace distribution causes a non-closed form formula. To solve this problem, it was assumed that the DCT coefficients were statistically independent, whereas DCT only

reduces the correlation between the coefficients, and they are not completely uncorrelated to each other. If we assume that the DCT coefficients are uncorrelated to each other and that the distribution of coefficients is Laplace, we cannot assume that the coefficients are statistically independent. If we use independent component analysis (ICA) instead of DCT, whose coefficients are greatly statistically independent from each other, we can do a more accurate statistical modeling.

We modeled the clean speech signal using HMM in ICA domain with Laplace distribution, while doing the noise modeling only by assuming Gaussian pdf for noises. In this work, we propose a novel MAP HMM-based speech enhancement algorithm that uses the ICA transformation. Our theoretical analysis shows that under the assumption of Laplace clean speech and Gaussian noise, the proposed algorithm leads to a well-known enhancement technique, sparse code shrinkage. This paper is organized as follows. In Section 2, the HMM training methods are reviewed. In Section 3, the MAP estimator is derived based on HMM in the ICA space. In Section 4, a summary of the proposed algorithm is given. In Section 5, we present the experimental evaluation and results, and in Section 6, the conclusions are given.

## 2. Signal model

Assume a time-domain noisy speech vector  $y_n$  at time  $n$  that is composed of a clean speech vector  $s_n$  and an additive noise vector  $d_n$  given as (1). Taking the independent component analysis (ICA) of  $y_n$ , we get (2) in this equation. We assumed that noise is independent from clean speech, and that the vectors have the length  $L$  and a zero mean. The AR features of  $P^{th}$  order  $a_n=[1, a(1), \dots, a(p)]$  for  $s_n=[s(0), s(1), \dots, s(L-1)]$  could be derived by the linear predictive coefficient approach [9], and the AR coefficients of other signals are obtained analogously.

$$y_n = s_n + d_n \tag{1}$$

$$Y_{ICA} = S_{ICA} + D_{ICA} \tag{2}$$

An HMM with  $M$  states and  $N$  mixtures is defined as  $\lambda = (\pi_{1 \times M}, a_{M \times M}, c_{M \times N}, \delta_{M \times N \times L})$ , where  $\pi$  is the initial state distribution,  $a$  denotes the state transition probability distribution,  $c$  is the probability distribution for each mixture in each state, and  $\delta_{M \times N \times L}$  is the matrix of pdf parameters in each mixture. The parameters of  $\lambda$  are estimated by the Baum re-estimation formulas

[10]. In order to estimate clean speech from noisy signal, it is necessary to construct the HMM models for clean speech ( $\lambda_s$ ) and noise ( $\lambda_d$ ) separately, and then combine them to create the noisy HMM ( $\lambda_y$ ).

### 2.1. Speech model

Based on the central limit theorem, we can assume that  $S_{ICA}$  has a multivariate Gaussian pdf with independent coefficients according to (3) and (5). In these equations, index  $k$  shows the  $k^{th}$  ICA coefficient of an  $L$ -dimensional vector.

As shown in [6], the Laplace distribution function, compared to the Gaussian distribution function, is closer to the speech signal distribution in different domains, and thus we can consider the distribution of vector  $S_{ICA}$  as a multivariate Laplace pdf. We know that the ICA coefficients are independent. Therefore, the multivariate Laplace pdf of  $S_{ICA}$  is derived by (4) and (5), where  $b_k$  is the scale parameter of the  $k^{th}$  coefficient. In these equations, it is assumed that  $S_{ICA}$  has a zero mean.

$$p(S_{ICA}(k)) = \frac{1}{\sqrt{2\pi}\sigma_k} \exp\left(-\frac{(S_{ICA}(k))^2}{2\sigma_k^2}\right) \tag{3}$$

$$p(S_{ICA}(k)) = \frac{1}{2b_k} \exp\left(-\frac{|S_{ICA}(k)|}{b_k}\right) \tag{4}$$

$$p(S_{ICA}) = \prod_{k=0}^{L-1} p(S_{ICA}(k)) \tag{5}$$

We used (4) and (5) for each mixture in each state, and estimated the model parameters of  $\lambda_s$  in closed form using the Baum's auxiliary function. In fact, changing Gaussian pdf to Laplace pdf in each HMM mixture modifies the equations for parameter estimation of  $\delta_s$  (the Laplace scale parameter estimation in each mixture in each state). Estimation of the Laplace scale parameter can be derived by differentiating the auxiliary function of (6) with respect to scale parameter resulting in (7).

$$Q(\lambda_s, \lambda_s^{old}) = \sum_{j=1}^M \gamma_j(j) \ln(\pi_j) + \sum_{n=1}^{T-1} \sum_{i=1}^M \sum_{j=1}^M \varepsilon_n(i, j) \ln(a_{ij}) \tag{6}$$

$$+ \sum_{n=1}^{T-1} \sum_{j=1}^M \gamma_n(j) \ln(p(S_{ICA} | \varphi_j))$$

$$\bar{b}_{j,k} = \frac{\partial Q}{\partial b_{j,k}} = \frac{\sum_{n=1}^{T-1} \gamma_n(j) |S_{ICA}|}{\sum_{n=1}^{T-1} \gamma_n(j)} \tag{7}$$

In the above-mentioned equations,  $\gamma_n(j)$  is the probability of being in state  $j$  at time  $n$ ,  $\varepsilon_n(i, j)$  is the transition probability of state  $i$  at time  $n$  to

state  $j$  at time  $n+1$ , and  $p(S_{ICA} | \varphi_j)$  is obtained using (4) and (5), where  $\varphi_j(k)$  equals the scale parameter of the  $k^{th}$  dimension in state  $j$  ( $\varphi_j(k) = b_{j,k}$ ).

**2.2. Noise model**

In this work, we assumed that  $D_{ICA}$  had a multivariate Gaussian distribution. In other words, if we use (3) and (5) for each mixture in each state, then we can estimate the model parameters of  $\lambda_D$  using the Baum’s auxiliary function. Therefore, estimation of  $\delta_{D,M \times N \times L}$  in model  $\lambda_D$ , can be interpreted as estimation of the diagonal covariance matrix ( $\Sigma_D$ ) for each mixture in each state. This means that the main diagonal of the covariance matrix contains the variances for each independent dimension.

**3. Map estimation**

In this section, we present the MAP estimation based on the Hidden Markov model that in the ICA space. We assumed, that the speech distribution was non-Gaussian, and that the noise distribution was Gaussian. Studies have shown that the proposed framework under the assumption that the signal is non-Gaussian and the noise is Gaussian leads to a sparse code shrinkage [11], which we called the SCS-HMM technique.

Let  $s_t$  be an  $L$ -dimensional vector of the clean speech. Similarly, let  $d_t$  be an  $L$ -dimensional vector of the noise. Assume that the noise is additive and statistically independent of the speech. Let  $y_t = s_t + d_t$  be an  $L$ -dimensional vector of the noisy speech. Let  $s_0^T = \{s_t; t = 1:T\}$ ,  $d_0^T = \{d_t; t = 1:T\}$ , and  $y_0^T = \{y_t; t = 1:T\}$ . The MAP estimation of clean speech  $s_t$ , given as  $y_0^T$ , is obtained by maximizing  $p(s_t | y_0^T)$  over  $s_t$ . Therefore, we applied the EM algorithm for the iterative local maximization of  $p(s_t | y_0^T)$ .

In this method, at each iteration, the auxiliary function is maximized in (8),

$$Q(s_t^{k+1}) = \sum_{q,u} p_\lambda(q,u | s_t^k) \ln p_\lambda(q,u, s_t^{k+1} | y_t) \quad (8)$$

where,  $s_t^k$  and  $s_t^{k+1}$  denote the estimate of  $s_t$  as obtained in the  $k^{th}$  and  $k+1^{th}$  iteration, respectively. Maximization of  $Q(s_t^{k+1})$  over  $s_t^{k+1}$  results in an estimate for which

$\ln p(s_t^{k+1} | y_t) \geq \ln p(s_t^k | y_t)$ , where equality holds if and only if  $s_t^{k+1} = s_t^k$ .

Maximization of (8) results in the following signal re-estimation formula:

$$\begin{aligned} s_t^{k+1} &= \arg \max_{s_t^{k+1}} \sum_{q,u} p_\lambda(q,u | s_t^k) \ln p_\lambda(q,u, s_t^{k+1} | y_t) \quad (9) \\ &= \sum_{q,u} p_\lambda(q,u | s_t^k) \arg \max_{s_t^{k+1}} (\ln p_\lambda(q,u, s_t^{k+1} | y_t)) \\ &= \sum_{q,u} p_\lambda(q,u | s_t^k) \arg \max_{s_t^{k+1}} (\ln p_\lambda(q,u, s_t^{k+1} | y_t)) \end{aligned}$$

where,

$$\begin{aligned} p_{\lambda_{\lambda_d}}(q,u, s_t^{k+1}, y_t) &= p_{\lambda_{\lambda_d}}(y_t | q,u, s_t^{k+1}) p_{\lambda_s}(q,u, s_t^{k+1}) \quad (10) \\ &= p_{\lambda_d}(y_t | s_t^{k+1}) p_{\lambda_s}(q,u, s_t^{k+1}) = p_{\lambda_d}(y_t - s_t^{k+1}) p_{\lambda_s}(q,u, s_t^{k+1}) \end{aligned}$$

On substituting (10) into (9), we obtained the following formula:

$$\begin{aligned} s_t^{k+1} &= \sum_{q,u} p_\lambda(q,u | s_t^k) \quad (11) \\ &\arg \max_{s_t^{k+1}} (\ln p_{\lambda_d}(y_t - s_t^{k+1}) + \ln p_{\lambda_s}(q,u, s_t^{k+1})) \end{aligned}$$

For a Gaussian distributed noise, the term  $\ln p_{\lambda_d}(y_t - s_t^{k+1})$  in (11) has no extremum. Therefore, the maximization in (11) is decided only by the term  $\ln p_{\lambda_s}(q,u, s_t^{k+1})$ . In order to estimate the clean signal in the ICA space, we used the ICA unmixing matrix  $w_s$ , obtained from the training phase. Thus, the estimate of signal  $s$  can be obtained by letting  $w_d = w_s$ . For clarity of presentation, we denoted  $w_s$  by  $w$ . In this case, the MAP estimation rule from (11) can be expressed in the form of:

$$\begin{aligned} s_t^{k+1} &= \sum_{q,u} p_\lambda(q,u | w, s_t^k) \arg \max_{s_t^{k+1}} (\sum_{l=1}^L (\ln p_{\lambda_d}(w_l.(y_t - s_t^{k+1})) \\ &+ \ln p_{\lambda_s}(q,u, w_l.(s_t^{k+1}))) + 2 \ln |\det(w)|) \quad (12) \end{aligned}$$

where,  $w_l = w(l,:)$  denotes the  $l^{th}$  row of matrix  $w$ . In (12), the conditional probability  $p_\lambda(q,u | w, s_t^k)$  is calculated by the forward-backward algorithm [12], and the second term of the above equation is calculated as follows:

$$\begin{aligned} s_t &= \arg \max_{s_t} (\sum_{l=1}^L \ln p_{\lambda_d}(w_l.(y_t - s_t)) \\ &+ \ln p_{\lambda_s}(q,u, w_l.(s_t))) + 2 \ln |\det(w)| \quad (13) \end{aligned}$$

We can perform the estimation in the independent space first, and then transform the estimate obtained into the original space. Denote  $w_l y_t$  as  $z_t(l)$  and  $w_l s_t$  as  $x_t(l)$ . Thus the components of

vector  $x_t$  can be calculated by the equations (14) and (15):

$$x_t = \begin{bmatrix} x_t(1) \\ \vdots \\ x_t(L) \end{bmatrix} \quad (14)$$

$$x_t(l) = \arg \max_x (\ln p_{\lambda_d}(z_t(l) - x_t(l)) + \ln p_{\lambda_s}(x_t(l) | q, u)) \quad (15)$$

$$\Leftrightarrow \arg \min_x \left( \frac{1}{2\sigma^2} (z_t(l) - x_t(l))^2 + f^{u|q}(x_t(l)) \right)$$

where,  $f^{u|q}(x_t(l)) = -\ln p_{\lambda_s}(x_t(l) | q, u)$ . The minimization is equivalent to solving the following equation:

$$\frac{1}{\sigma^2} (x_t(l) - z_t(l)) + f^{u|q}(x_t(l)) = 0 \quad (16)$$

Although (16) may not have a closed form solution, the estimation function can be approximated as follows [13]:

$$x_t(l) = \text{sign}(z_t(l)) \max(0, |z_t(l)| - \sigma^2 \frac{1}{b_{u|q}}) \quad (17)$$

In the above equation,  $b_{u|q}$  is the scale parameter of the fourth mixture in the  $q$ th state. The estimation rule in (17) is known as the sparse code shrinkage estimation [11]. Given the two words  $p_{\lambda}(q, u | w, s_t^k)$  and

$$\sum_{l=1}^L \arg \max_x (\ln p_{\lambda_d}(w_l(y_t - s_t)) + \ln p_{\lambda_s}(w_l(s_t) | q, u)),$$

we can estimate the clean signal component by (11).

#### 4. Summary of the proposed SCS-HMM algorithm

This section provides a summary of the steps involved in the proposed SCS-HMM enhancement algorithm, as described in Sections 2 and 3.

1. First, using two sets of data  $\tilde{s}$  and  $\tilde{d}$ , which should have the same statistical properties as the noise  $d$  and signal  $s$ , calculate the ICA transformation matrices  $w_d$  and  $w_s$ . This can be performed using any of the existing ICA algorithms.
2. Train HMM using the independent components  $S_{ICA} = w_s \tilde{s}$  and  $D_{ICA} = w_d \tilde{d}$ , as described in Sections 2.1 and 2.2, respectively.

3. To perform the enhancement process of the observed noisy signal  $y$ , we applied the estimation rule (11) to estimate the clean speech  $\hat{s}$ .

#### 5. Experimental evaluation

The objective evaluation of AR-HMM, LaGa-HMM, and the proposed algorithm was performed in terms of the SNR and PESQ measures. The experimental evaluation was performed using the speech signals selected from the TIMIT database, separately for each gender. The training set contained 100 sentences for each gender, and the testing set contained the sentences of the speaker female and male. There were no common sentences between the training and test sets. The noisy speech signals were created by adding different noises such as the white, babble, and machinegun noises at 0 db, 5 db, and 10 db. All of the signals were sampled at 8 kHz. The signals were split into frames of 64 samples using the rectangular window. The fast-ICA algorithm [14] was employed to estimate the ICA basis functions based on the training set. In the training phases of the various models, there was no inter-frame overlap. The clean speech models were generated using 10 states and 30 mixtures. In AR-HMM and LaGa-HMM, the noise models were constructed based on 4 states and 4 mixtures. Due to the use of MAP estimation in the SCS-HMM method, the noise model was generated using 1 state and 1 mixture. In AR-HMM, we used an AR-order of 10 for a clean speech and noise. The fast-ICA algorithm was employed to estimate the ICA basis functions based on the training data. The performance of the proposed algorithm SCS-HMM was compared with AR-HMM in [12], and LaGa-HMM<sub>DCT</sub> in [8]. Tables 1 and 2, respectively, demonstrate the SNR and PESQ values achieved by AR-HMM, LaGa-HMM<sub>DCT</sub>, and SCS-HMM. The results obtained confirm the superiority of the SCS-HMM method in the presence of non-stationary noises compared to LaGa-HMM<sub>DCT</sub> based on the SNR measure. The results of the SCS-HMM method represent a better performance of this method, compared to AR-HMM, in the presence of white noise based on the PESQ measure.

This result is expected, because the MAP method is constructed based on stationary noises such as the white noise. For this reason, the performance of the SCS-HMM method is better than AR-HMM in the presence of white noise based on the PESQ measure. However, improvements in SNR results were not observed. An example of clean, noisy, and enhanced speech spectrograms are depicted in figure 1. Using the spectrogram

representation, it can be seen that the SCS-HMM result is close to the original clean signal.

**Table 1. Comparative performance in terms of SNR.**

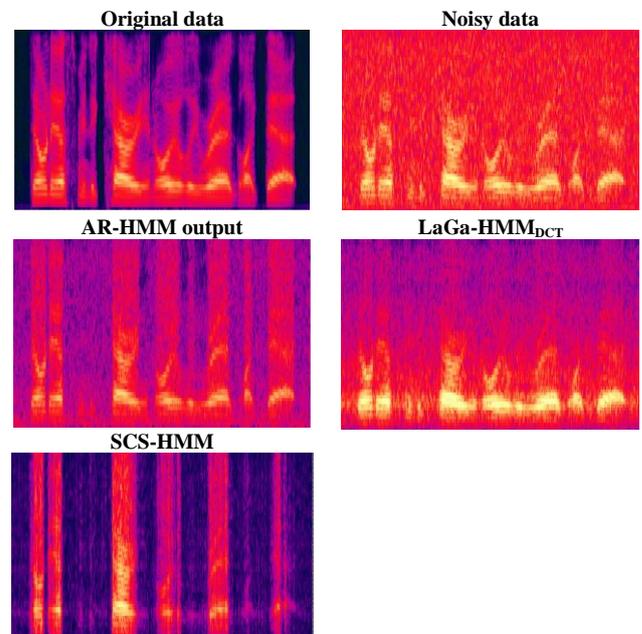
Enhancement algorithm		Babble noise			White noise			Machinegun noise		
		0dB	5dB	10dB	0dB	5dB	10dB	0dB	5dB	10dB
Male	AR-HMM	8.73	11.34	14.53	5.42	8.60	11.56	8.73	11.34	14.23
	LaGa-HMM <sub>DCT</sub>	7.96	10.55	13.67	4.24	7.17	10.70	7.63	9.08	10.93
	SCS-HMM	8.59	10.89	13.79	5.34	8.34	11.49	7.36	9.24	11.53
Female	AR-HMM	6.37	9.44	12.87	7.29	9.74	12.43	10.35	13.34	16.14
	LaGa-HMM <sub>DCT</sub>	5.41	7.75	11.57	6.93	9.06	11.16	7.92	9.63	12.10
	SCS-HMM	6.22	8.42	12.20	7.22	9.56	12.03	8.91	10.29	12.58

**Table 2. Comparative performance in terms of PESQ.**

Enhancement algorithm		Babble noise			White noise			Machinegun noise		
		0dB	5dB	10dB	0dB	5dB	10dB	0dB	5dB	10dB
Male	AR-HMM	2.05	2.41	2.73	1.79	2.08	2.33	2.80	3.01	3.25
	LaGa-HMM <sub>DCT</sub>	1.75	2.24	2.59	1.76	2.03	2.30	2.62	2.93	3.08
	SCS-HMM	1.92	2.29	2.70	2.01	2.20	2.41	2.49	2.78	3.05
Female	AR-HMM	1.86	2.13	2.46	1.66	1.90	2.15	2.72	3.02	3.31
	LaGa-HMM <sub>DCT</sub>	1.64	2.08	2.38	1.56	1.85	2.09	2.60	2.85	3.23
	SCS-HMM	1.71	2.10	2.44	1.70	1.90	2.23	2.47	2.80	3.14

**6. Conclusion**

In this paper, we presented a new HMM-based speech enhancement framework based on the independent component analysis (ICA). Furthermore, a MAP estimator was derived for the ICA coefficients of a clean speech. It was also shown that the proposed framework under the assumption of the signal being Laplace distribution and noise being Gaussian distribution led to sparse code shrinkage, called the SCS-HMM technique. The evaluation results, in terms of SNR and PESQ, indicated the superiority of the SCS-HMM method in the presence of non-stationary noises, compared to LaGa-HMM<sub>DCT</sub>. The results of the SCS-HMM method represented a better performance of this measure. The performance of SCS-HMM in the presence of other noise types based on PESQ and in the presence of all noises based on SNR showed slightly inferior performance of this method.



**Figure 1. Spectrograms of female speech corrupted by white noise at SNR=5 dB.**

## References

- [1] Zhao, D. Y. & Kleijn, W. B. (2007). HMM-Based Gain Modeling for Enhancement of Speech in Noise. *IEEE Transactions on Audio, Speech and Language Processing*, vol. 15, pp. 882-892.
- [2] Sameti, H. & Deng, L. (2002). Nonstationary-state hidden Markov model representation of speech signals for speech enhancement. *Signal Process.*, vol. 82, pp. 205-227.
- [3] Srinivasan, S., Samuelsson, J. & Kleijn, W. B. (2006). Codebook Driven Short-Term Predictor Parameter Estimation for Speech Enhancement. *IEEE Trans. on Audio, Speech & Language Processing*, vol. 14, no. 1, pp.163-176.
- [4] Martin, R. (2005). Speech Enhancement Based on Minimum Mean-Square Error Estimation and Super-Gaussian Priors. *IEEE Trans. on Speech Audio Proc.*, vol. 13, no. 5, pp. 845–856.
- [5] Lotter, T. & Vary, P. (2005). Speech Enhancement by MAP Spectral Amplitude Estimation Using a Super-Gaussian Speech Model. *EURASIP Journal Appl. Signal Proc.*, vol. 2005, pp. 1110–1126.
- [6] Gazor, S. & Zhang, W. (2003). Speech probability distribution. *IEEE Signal Processing Letters*, vol. 10, no. 7, pp. 204-207.
- [7] Chen, B. & Loizou, P. (2007). A Laplacian-Based MMSE Estimator for Speech Enhancement. *Speech Communication*, vol. 49, no. 2, pp. 134-143.
- [8] Aroudi, A., Veisi, H. & Sameti, H. (2012). Speech Enhancement Based on Hidden Markov Model with Discrete Cosine Transform Coefficients Using Laplace and Gaussian Distributions. in *ISSPA, Conference on Information Science, Signal Processing and their Application*, Montreal, Canada, pp 304-309.
- [9] MARKEL, J. D., & GRAY, A. H. (2013). *Linear Prediction of Speech*, Springer-Verlag, New York.
- [10] Bishop, C. M. (2006). *Pattern Recognition and Machine Learning*, Springer-Verlag New York.
- [11] Wipf, D. (2006). Bayesian methods for finding sparse representations, Ph.D. Thesis, University of California, San Diego.
- [12] Ephraim, Y., Malah, D. & Juang, B. H. (1989). On the Application of Hidden Markov Models for Enhancing Noisy Speech. *IEEE Transactions on Acoustics, Speech, and Signal Processing*, Vol. 37, No. 12, pp. 1846-56.
- [13] Zou, X., Jancovic, P., Liu, J. & Köküer, M. (2008). Speech Signal Enhancement Based on MAP Algorithm in the ICA Space. *IEEE Transactions on Signal Processing*, vol. 56, no. 5, pp. 1812-1820.
- [14] Langlois, D., Chartier, S. & Gosselin, D. (2010). independent component analysis: InfoMax and FastICA algorithms. *Tutorials in Quantitative Methods for Psychology*, vol. 6, no. 1, pp. 31–38.

## بهسازی گفتار بر پایه مدل مخفی مارکوف با استفاده از گرفتگی کدهای تنک

الهام گلرسان\* و حسین صامتی

دانشکده مهندسی کامپیوتر، دانشگاه صنعتی شریف، تهران، ایران

ارسال ۲۰۱۵/۰۸/۱۹؛ پذیرش ۲۰۱۶/۰۵/۳۱

### چکیده:

این مقاله یک چارچوب جدید برای بهسازی گفتار با مدل مخفی مارکوف و مبتنی بر تحلیل مؤلفه مستقل (ICA) ارائه می‌کند. مراحل تحلیل برای آموزش مدل‌های گفتار و نویز با استفاده از روش بازتخمین باوم ارائه می‌شود و یک تخمین گر MAP مبتنی بر ترکیب توزیع لاپلاس-گوسی (به ترتیب برای سیگنال‌های گفتار تمیز و نویز) در ساختار مبتنی بر مدل مخفی مارکوف تحت عنوان SCS-HMM پیشنهاد می‌شود. روش پیشنهادی روی دادگان TIMIT با سه نوع نویز در سه مقدار SNR ورودی با معیارهای PESQ و SNR خروجی مورد ارزیابی قرار می‌گیرد و با دو روش AR-HMM و نیز روش بهسازی برپایه ضرایب تبدیل کسینوسی گسسته با استفاده از توزیع‌های لاپلاس و گوسی (LaGa-HMM<sub>DCT</sub>) مقایسه می‌شود. نتایج به دست آمده نشان‌دهنده برتری SCS-HMM در حضور نویزهای نالیستان نسبت به LaGa-HMM<sub>DCT</sub> است. همچنین نتایج SCS-HMM عملکرد بهتر این روش نسبت به AR-HMM را در حضور نویز سفید با معیار PESQ نشان می‌دهد.

**کلمات کلیدی:** بهسازی گفتار، توزیع لاپلاس چندمتغیره، تحلیل مولفه مستقل، بهسازی گفتار مبتنی بر مدل مخفی مارکوف