Advanced Science and Technology Letters
Vol.30 (ICCA 2013), pp.159-163
http://dx.doi.org/10.14257/astl.2013.30.34

Adaptive Noise Estimation Based on Non-negative Matrix Factorization

Kwang Myung Jeon and Hong Kook Kim

School of Information and Communications
Gwangju Institute of Science and Technology (GIST)
{kmjeon, hongkook}@gist.ac.kr

Abstract. In this paper, an adaptive noise estimation technique is proposed on the basis of non-negative matrix factorization (NMF). As an initial step of the proposed method, the noise basis matrix of NMF is estimated from a collection of noise signals. Then, the proposed method updates the initially estimated noise basis matrix on the fly by using an estimate of the noise spectrum from the noisy signal. It is here demonstrated that the proposed method provides a better noise estimate than a NMF-based method without using any adaptation, especially when there is a mismatch in noise conditions for noise basis training and estimation using NMF.

Keywords: Noise estimation, non-negative matrix factorization (NMF), mismatched noise condition, basis adaptation

1 Introduction

In general, noise estimation methods based on signal-to-noise ratio (SNR), such as Wiener filtering [1] or minimum mean squared error log-spectral amplitude (MMSE-LSA) [2], work well under stationary noise conditions. However, they may not accurately estimate non-stationary noises that occur in most real environments [3]. As an alternative, non-negative matrix factorization (NMF) based noise reduction methods have been proposed [4][5] to effectively estimate noise spectrum under non-stationary noise conditions. Nevertheless, the performance of the NMF-based noise estimation method can be limited depending on how accurately the noise basis matrix can be used to decompose a noisy signal into a clean and noise signals.

In this paper, an adaptive noise estimation method is proposed in an NMF framework. As an initial step of the proposed method, the noise basis matrix of NMF is estimated from a collection of noise signals. Then, the proposed method updates the initially estimated noise basis matrix on the fly by using an estimate of noise spectrum from the noisy signal. Thus, the noise spectrum is estimated from the adapted noise basis matrix in the proposed method, while conventional NMF-based noise estimation methods rely on prior knowledge of a certain type of noise. Therefore, it is expected that the proposed method should be able to more accurately estimate noise spectrum than the conventional methods, especially when there is a mismatch in noise conditions for noise basis training and estimation using NMF.

ISSN: 2287-1233 ASTL

Following this introduction, Section 2 proposes an NMF-based adaptive noise estimation method. Subsequently, Section 3 evaluates the performance of the proposed noise estimation method. Finally, Section 4 concludes this paper.

2 Proposed NMF-Based Adaptive Noise Estimation Method

Let $y_i(n)$, $s_i(n)$, and $d_i(n)$ be noisy speech, clean speech, and additive noise for the i-th analysis frame, respectively, where $d_i(n)$ is uncorrelated with $s_i(n)$. By applying a short-time Fourier transform (STFT), $y_i(n)$ can be represented as spectral components, such as $Y_i k\ S_i k\ D_i k$

$( ) = ( ) + ( )$ for $k = 0, \cdots, K$ 1, where $Y_i(k)$, $S_i(k)$, and $D_i(k)$ denote the k-th spectral components of $y_i(n)$, $s_i(n)$, and $d_i(n)$, respectively. The goal of the proposed noise estimation method is to estimate $D_i(k)$ from $Y_i(k)$ without prior knowledge of $S_i(k)$ and $D_i(k)$. To this end, several analysis frames are concatenated so that $\mathbf{Y} = \mathbf{S} + \mathbf{D}$ is obtained. Thus, all the matrices, $\mathbf{Y}$, $\mathbf{S}$, and $\mathbf{D}$, are all $K \cdot N$ matrices, where K and N are the number of frequency bins and the number of concatenated frames, respectively.

In the NMF framework, $\mathbf{D}$ is represented as a form of $\mathbf{D} = \mathbf{B}_D\mathbf{A}_D$, where $\mathbf{B}_D$ and $\mathbf{A}_D$ are a basis matrix and an activation matrix of $\mathbf{D}$, respectively. First of all, the NMF training [5] is applied to obtain an initial noise basis matrix, $\mathbf{B}_D$, from the collection of noise signals. After that, the noise basis adaptation is performed iteratively by the following equations

$$\mathbf{B}_D^{t+1} = \mathbf{B}_D^t \otimes \frac{\hat{\mathbf{D}}}{\mathbf{B}_D^t \mathbf{A}_D^t \mathbf{1}} \mathbf{A}_D^{t\,T}$$

$$\mathbf{A}_D^{t+1} = \mathbf{A}_D^t \otimes \frac{\mathbf{B}_D^{t\,T} \hat{\mathbf{D}}}{\mathbf{B}_D^{t\,T} \mathbf{B}_D^t \mathbf{A}_D^t} \qquad (2)$$

where $\mathbf{B}_D^0 = \mathbf{B}_D$, t is an iteration index, T is the transpose operator, and $\mathbf{1}$ is a $K \cdot N$ matrix with all elements equal to unity. Moreover, both multiplication, $\otimes$, and division indicate element-wise operators. In Eqs. (1) and (2), $\hat{\mathbf{D}}$ is a spectral magnitude matrix of initial noise signal within short-pause region of noisy speech signal. Note that the duration of the short-pause region is set to 0.5 sec in this paper. Next, $\mathbf{B}_D$ and $\mathbf{A}_D$ are a $K \cdot a$ basis matrix and an $a \cdot N$ activation matrix obtained after the t-th iteration, respectively, where a is the number of bases and it is a controllable parameter in NMF. Note that all elements for $\mathbf{A}_D$ are set as random values between 0 and 1, and $\mathbf{B}_D^0 = \mathbf{B}_D$. The iterative procedure described above is

terminated if the difference of an objective function according to the iteration is less than a pre-defined threshold. That is, the objective function is defined as [7]

$$obj(t) = \sum_{K,N} \left[\left(\mathbf{D} \log \frac{\mathbf{D}}{\mathbf{A}_6^t} \right) \mathbf{B}_n^t \mathbf{A}_n^t \right] \quad (3)$$

where summation means the addition of all the elements of a matrix. Accordingly, if $|obj(t) - obj(t-1)| / obj(t-1) < \theta$, then the procedure described in Eqs. (1) and (2) is terminated, where θ is manually set to 100 in this paper by trading off the iteration number and the adaptation accuracy. Consequently, $\mathbf{B}_D = \mathbf{V}_D$ is obtained if the procedure is terminated at the t-th iteration.

Next, **S** is estimated, which corresponds to finding $\mathbf{A}_D$ from **Y** by applying the equations of

$$\mathbf{B}_S^t = \mathbf{B}_S^{t-1} \square \frac{\mathbf{Y}}{\mathbf{B}_S^{t-1} \mathbf{A}_S^{t-1}} \mathbf{A}_S^{T}, \quad (4)$$

$$\mathbf{A}_Y^t = \mathbf{A}_Y^{t-1} \square \frac{\mathbf{B}_Y^{t-1} \mathbf{Y}}{\mathbf{B}_Y^{t-1} \mathbf{A}_Y^{t-1}} \quad (5)$$

where $\mathbf{B}_S$ and $\mathbf{A}_S$ are an $K \times a$ basis matrix and an $a \times N$ activation matrix of **S**, respectively. Moreover, $\mathbf{B}_Y = [\mathbf{B}_S ; \mathbf{B}_D]$ and $\mathbf{A}_Y = [\mathbf{A}_S ; \mathbf{A}_D]$. Note that all elements of $\mathbf{B}_S$ and $\mathbf{A}_Y^0$ are set as random values between 0 and 1. Similarly to the termination condition in Eq. (3), the estimation procedure is finished by checking whether $\left[\sum_{K,N} \left(\mathbf{Y} \log(\mathbf{Y} / \mathbf{B}_Y^t \mathbf{A}_Y^t) \right) \left(\mathbf{B}_Y \mathbf{A}_Y \right) \right]$ is going to be converged. Finally, if the procedure of Eqs. (4) and (5) is terminated at the t-th iteration, the noise spectrum, **D**, is estimated as $\mathbf{D} = \mathbf{B}_D^t \cdot \mathbf{A}_D^t$.

3 Performance Evaluation

The performance of a noise estimation method was evaluated by measuring the log-spectral distance (LSD) in dB [8] between the true noise spectrum and the estimated one by the noise estimation method. To this end, 10 speakers uttered 20 sentences each, resulting in 200 sentences in total. Each sentence was mixed with home TV noise at around 20 dB SNR. Note here that home TV noise was used to simulate a highly non-stationary noisy environment in which a person was talking while watch-

ing different genres of TV programs, such as dramas, news, sports, and movies. In addition, in order to simulate mismatched noise condition, $\mathbf{B}_{\sim \mathbf{D}}$, for Eqs. (1) and (2),

Table 1. Performance comparison of the noise estimation methods in LSD (dB).

Speaker	Conventional	Proposed	LSD reduction
A	2.11	0.54	1.57
B	1.87	0.61	1.26
C	1.34	0.83	0.51
D	1.53	0.55	0.98
E	1.76	0.67	1.09
F	1.87	0.61	1.26
G	2.01	0.49	1.52
H	1.75	0.57	1.18
I	1.35	0.46	0.89
J	2.04	0.61	1.43
Average	1.76	0.59	1.17

was obtained from 10 minutes of noise signal consisting of bus stops, restaurants, and subway noises. Moreover, K, a , and N were set to 257, 40, 300, respectively.

Table 1 compares the LSDs between the proposed method and a conventional NMF-based method that did not have any adaptation scheme for the noise estimation [5]. That is, the conventional method used $\mathbf{B}_{\mathbf{Y}}[\mathbf{B}_{S}; \mathbf{B}_{D}]$ in Eq. (4) instead of $\mathbf{B}_{\mathbf{Y}}[\mathbf{B}_{S}; \mathbf{B}_{D}]$.

It was shown from the table that the proposed method reduced average LSD by 1.17 dB, compared to the conventional method. This implies that the proposed method could provide a better noise spectral estimate than the conventional method.

4 Conclusion

In this paper, an NMF-based adaptive noise estimation method was proposed to overcome a mismatch in noise conditions for noise basis training and estimation using NMF. The proposed method first performed the noise basis adaptation to update the initially estimated noise basis matrix on the fly by using an estimate of noise spectrum from the noisy signal. Then, noise spectrum of current noise environment was estimated by NMF with the adaptive noise basis matrix. It has been shown from the experiment that the proposed method provided lower LSD than the conventional NMF-based method.

Acknowledgments. This work was supported in part by the IT R&D program of MSIP/KEIT [10035252, Development of dialog-based spontaneous speech interface technology on mobile platform] and the National Research Foundation of Korea (NRF) grant, funded by the government of Korea (MSIP) (No. 2012-010636).

References

1. Lim, J., Oppenheim, A. V.: All-pole modeling of degraded speech. IEEE Transactions on Acoustics, Speech, and Signal Processing, 26(3), (1978) pp. 197–210.
2. Ephraim, Y., Malah, D.: Speech enhancement using a minimum mean-square error short-time spectral amplitude estimator. IEEE Transactions on Acoustics, Speech, and Signal Processing, 32(6), (1984) pp. 1109–1121.
3. Rangachari, S., Loizou, P., Hu, Y.: A noise estimation algorithm with rapid adaptation for highly nonstationary environments. In: Proceedings of ICASSP, (2004) pp. 305–308.
4. Kim, S. M., Park, J. H., Kim, H. K., Lee, S. J., Lee, Y. K.: Non-negative matrix factorization based noise reduction for noise robust automatic speech recognition. Lecture Notes in Computer Science, 7191, (2012) pp. 338–346.
5. Jeon, K. M., Park, N. I., Kim, H. K., Hwang, K. I., Choi, M. K.: Non-stationary noise estimation based on non-negative matrix factorization. Advanced Science and Technology Letters, 14, (2012) pp. 172–175.
6. Lee, D. D., Seung, H. S.: Learning the parts of objects by non-negative matrix factorization. Nature, 401(6755), (1999) pp. 788–791.
7. Lee, D. D., Seung, H. S.: Algorithms for nonnegative matrix factorization. Advances in Neural Information Processing Systems (NIPS), 13, (2000) pp. 556–562.
8. Gray, A. H., Markel, J. D.: Distance measures for speech processing. IEEE Transactions on Acoustics, Speech, and Signal Processing, 24(5), (1976) pp. 380–391.

