A SIGNAL SEPARATION METHOD BASED ON SPARSITY ESTIMATION OF SOURCE SIGNALS AND NON-NEGATIVE MATRIX FACTORIZATION

[1]SIYEON NAM, [2]SERIN HONG, [3]DEOKGYU YUN, [4] SEUNG HO CHOI

[1,2,4] Dept. of Electronic and IT Media Engineering, Seoul National University of Science and Technology
[3] Dept. of Electronic Engineering, Seoul National University of Science and Technology
E-mail: [1]siyeon1850@naver.com, [2]srin129@hanmail.net, [3] deokkyuyun@gmail.com, [4]shchoi@snut.ac.kr

Abstract - In order to improve the signal separation performance of the non-negative matrix factorization (NMF), sparse non-negative matrix factorization (Sparse NMF, SNMF) was developed. Existing SNMF algorithm uses arbitrarily determined sparseness without considering the sparseness of individual sound sources. In this paper, we propose a new signal separation method that estimates the sparseness according to the characteristics of a sound source and applies it to SNMF algorithm. Experimental results show that the proposed method has better performance than the existing NMF and SNMF.

Keywords - Signal separation, Non-negative matrix factorization, Sparseness, Denoising.

I. INTRODUCTION

Recently, the need for signal separation method is increasing in the fields of speech recognition, speech enhancement, virtual reality (VR) sound reconstruction, etc.Non-negative matrix factorization(NMF) has been studied for signal separation that separates each signal from a mixed signal.In speech signal processing, NMF separates the spectrogram of a mixed signal into a basis matrix and an activation matrix [1].In order to improve the performance of existing NMF, sparse non-negative matrix factorization(SNMF) has been proposed [2].However, the SNMF does not reflect the sparseness of each sound source to be separated but arbitrarily chooses sparseness, and there is a limit in obtaining the optimized basis and activation matrices. The proposed SNMF applies the sparseness of speech and noise differently according to the characteristics of the sources, and the sparseness is determined by the activation matrix of each source signal.

II. EXISTING NMF AND SNMF

NMF decomposes a non-negative matrix into a product of the basis matrix and the activation matrix. Existing NMF algorithm aims at the following equation when W and H are the basis matrix and the activation matrix, respectively.

$$V \approx WH \quad (1)$$

The optimal W and H inEq. (1) are generally obtained by the following Eq. (2) that minimizes the Kullback-Leibler divergence (KL divergence)distance [3]. The difference between NMF and SNMFis whether the sparsenessμterm is absent or not.

$$\underset{W,H \geq 0}{\operatorname{argmin}} D(V \| WH) + \mu \| H \| \quad (2)$$

The $W$and H are updated by multiplicative update rule [4] as in Eq. (3) and $\widetilde{W}$represents the normalized W.

$$H \leftarrow H\frac{W^T(V/WH)}{W^T 1+\mu}, W \leftarrow W\frac{V/WH+\widetilde{W}\,\widetilde{W}^T}{1+\widetilde{W}\widetilde{W}^T(V/WH)} \quad (3)$$

In case of existing SNMF, the$\mu$is arbitrarily set.

III. PROPOSED SNMF

The proposed SNMF obtainsthe initial matrices W and H for each speech signal and noise signal using NMF as in Fig. 1.Each sparseness λ is computed as Eq. (4) where N is the total number of elements of activation matrix H.

$$\lambda = \frac{\sqrt{N}-\sum\sum|h_{ij}|/\sqrt{\sum\sum h_{ij}^2}}{\sqrt{N}-1} \quad (4)$$

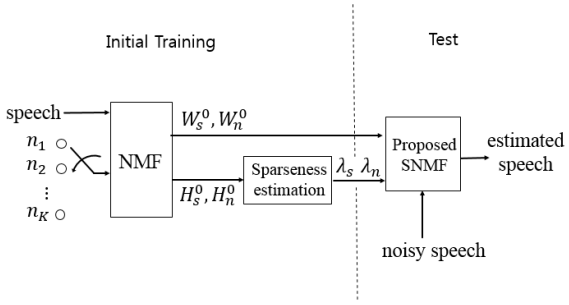

Figure1. Block diagram of proposed SNMF

Then, we construct a sparseness vector $\vec{\lambda} = [\lambda_s\ \lambda_s\ \ldots \lambda_s\ \lambda_n\ \lambda_n \ldots \lambda_n]^T$ usingthe speech sparsenessλ_s and noise sparsenessλ_n.In this study, we modify Eq. (3) by applying the vector$\vec{\lambda}$as Eq. (5).

$$H \leftarrow H\frac{W^T(V/WH)}{W^T 1(1+\vec{\lambda})}, W \leftarrow W\frac{V/WH+\widetilde{W}\,\widetilde{W}^T}{1+\widetilde{W}\widetilde{W}^T(V/WH)} \quad (5)$$

As a result, an estimated speech signal is obtained through the proposed SNMF using the sparseness $\vec{\lambda}$ and the initial matrices W_s^0, W_n^0 as in Fig. 1.

IV. EXPERIMENTANDRESULT

In order to compare the denoising performance of the proposed SNMF withthose of the existing NMF and SNMF, we used 1500 utterances of TIMIT DB [5] and 4 noises for initial training.In test, various noisy signalswere constructed using 100 speechutterances and 4 noisesdifferent from the database used for training.Performance was evaluated in terms of PESQ (Perceptual Evaluation of Speech Quality) [6]. First, we chose the μ value among{1, 5, 10, 15}, which gave the best PESQ score in existing SNMF.As shown in Table 1, it can be seen that theaverage PESQ scoreof the proposed SNMF is higher than those of existing NMF and SNMF.

Method / noise	Noisy	NMF	SNMF	Proposed SNMF
white	1.50	1.69	1.76	2.05
siren	1.99	2.10	2.02	2.27
car	3.15	3.29	3.44	3.40
factory	1.65	1.68	1.81	1.88
average	2.07	2.19	2.26	2.40

Table 1. Result of denoising performanceexperiment (PESQ)

CONCLUSION

In this paper, we proposed a novel SNMF-based signal separation method that applies the sparseness of speech and noise to SNMF algorithm. Experimental result showed that the denoising performance of the proposed SNMF is superior to those of the existingNMF and SNMF methods.

ACKNOWLEDGMENTS

This work was supported by the ICT R&D program of MSIP/IITP [2016-0-00144, Moving Free-viewpoint 360VR Immersive Media System Design and Component Technologies].

RFERENCES

[1] C. Joder, F. Weninger, and D. Schuller, “A comparative study on sparsity penalties for NMF-based speech separation: Beyond LP-norms,” IEEE International Conference. Acoustics, Speech and Signal Processing (ICASSP), pp. 858-862, 2013.

[2] J. Le Roux, F. Weninger, and J. R. Hershey, “Sparse NMF: half-baked or well done?,” Mitsubishi Electric Research Labs (MERL), Cambridge, MA, USA, Tech. Rep., no. TR2015-023, 2015.

[3] M. N. Schmidt, and R. K. Olsson, “Single-channel speech separation using sparse non-negative matrix factorization,” International Conference on Spoken Language Processing, 2006.

[4] P. Smaragdis, C. Fevotte, G. Mysore, N. Mohammadiha, and M. Hoffman, “Static and dynamic source separation using nonnegative factorizations: A unified view,” IEEE Signal Processing Magazine, vol. 31, no. 3, pp. 66-75, 2014.

[5] J. S. Garofolo, L. F. Lamel, W. M. Fisher, J. G. Fiscus, and D. S. Pallett, “DARPA TIMIT acoustic phonetic continuous speech corpus CD-ROM,” NTIS, 1993.

[6] RECOMMENDATION, ITU-T. “Perceptual evaluation of speech quality (PESQ): An objective method end-to-end speech quality assessment of narrow-band telephone networks and speech codecs,” Rec. ITU-T p.862, 2001.

★★★