Efficiency evaluation of subspace-based spectral subtraction based on iterative eigenvalue analysis in real environments

Yuki NAGANO[1]; Takahiro FUKUMORI[1];
Masato NAKAYAMA[2]; Takanobu NISHIURA[2]

[1] Graduate School of Information Science and Engineering, Ritsumeikan University, JAPAN

[2] College of Information Science and Engineering, Ritsumeikan University, JAPAN

ABSTRACT

In real environments, the recorded speech signal is much affected by unwanted noise. Therefore, it is necessary to reduce the unwanted noise from the recorded noisy signal. The spectral subtraction (SS) and the flooring processing-improved SS (F-SS) have been proposed to achieve that. The F-SS iteratively estimates the clean speech signal by utilizing the SS. However, the F-SS generates the distortion in the noise-reduced signal although it can reduce the unwanted noise. In this paper, we propose the subspace-based spectral subtraction (S-SS) to reduce the distortion from the noise-reduced signal. The proposed S-SS performs the eigenvalue analysis with multiple noise-reduced signals by the SS. The proposed S-SS acquires multiple noise-reduced signals by the SS under the various conditions of noise estimation. The subspace of speech component is calculated from multiple noise-reduced signals by the eigenvalue analysis. The proposed S-SS then acquires the noise-reduced signal which is reduced the distortion by using the subspace of the speech component. The proposed S-SS can simultaneously reduce the unwanted noise and the distortion of the observed signal by iteratively performing these processes. As a result of objective experiments with signal-to-distortion ratio (SDR), we confirmed that the proposed S-SS can reduce the distortion of the noise-reduced signal.

Keywords: Noise reduction, Spectral subtraction, Subspace
I-INCE Classification of Subjects Number(s): 01.4

1. INTRODUCTION

High quality speech recording is required for smart phones and video cameras. In real noisy environments, it is difficult that we record the speech signal without mixing the unwanted noise such as the PC fan noise and the factory noise. It is therefore necessary to reduce the unwanted noise from the observed signal.

The noise-reduced methods using a microphone array have been proposed (1, 2). These methods however require a large-scale system and high computing cost to reduce the unwanted noise in high accuracy. Also, the spectral subtraction (SS) (3) with a single microphone has been proposed. The SS reduces the unwanted noise from the observed signal by switching subtraction process and flooring process in the frequency domain. In addition, the SS estimates the unwanted noise from the non-speech segment of the observed signal. It can effectively reduce the unwanted noise at small-scale system and low computing cost. However, there is a problem that the SS generates a distortion and a residual noise called musical tone (4) in the noise-reduced signal. The flooring processing-improved spectral subtraction (F-SS) (5) has been proposed to reduce the musical tone. The musical tone is generated when the power differences between the subtracting part and the flooring part is generated in the noise-reduced signal. To reduce these power differences, the F-SS iteratively reduce a small amount of noise from the observed signal. It is thus possible to reduce the musical tone, but it is still difficult to reduce the distortion of the noise-reduced signal.

[1] {is0081ii, cm013061}@ed.ritsumei.ac.jp
[2] {mnaka@fc, nishiura@is}.ritsumei.ac.jp

In this paper, we propose the subspace-based spectral subtraction (S-SS) based on the iteratively SS and the iteratively eigenvalue analysis. In the proposed S-SS, we focus on that the different noise-reduced signal by the SS is generated by using the different noise segment and the different parameter of the noise estimation. The proposed S-SS acquires multiple noise-reduced signals by the SS under the various conditions of noise estimation. These noise-reduced signals include a common speech component, a different remnant noise component and a different distortion component. To reduce the distortion, we consider that only extracting the common speech component is effective. The proposed S-SS accordingly reduces the distortion of the noise-reduced signal by the eigenvalue analysis with these noise-reduced signals by the SS.

We finally objectively evaluate the effectiveness of the proposed S-SS in real noisy environments.

2. THE CONVENTIONAL METHOD

2.1 Spectral Subtraction (SS)

The SS (3) is the noise reduction method utilizing a simple algorithm. The SS reduces the unwanted noise from the observed signal in the frequency domain. The unwanted noise is estimated from the non-speech segment of the observed signal.

In the SS, the power spectrum of the noise-reduced signal is calculated by subtracting that of estimated noise from that of the observed signal. The SS further performs flooring processing when the power spectrum of the noise-reduced signal becomes negative value.

Equation (1) is the process of the SS.

$$|\hat{X}(\omega)|^2 = \begin{cases} |Y(\omega)|^2 - \alpha|\hat{N}(\omega)|^2, & \text{if } |Y(\omega)|^2 - \alpha|\hat{N}(\omega)|^2 > 0, \\ \beta|Y(\omega)|^2, & \text{otherwise,} \end{cases} \tag{1}$$

where $|\hat{X}(\omega)|^2$, $|\hat{N}(\omega)|^2$ and $|Y(\omega)|^2$ indicate the power spectrum of the noise-reduced signal, the estimated noise and the observed signal, respectively, α indicates subtraction coefficient and β indicates flooring coefficient. The subtraction coefficient is generally defined as $\alpha > 1.0$. The flooring coefficient is generally defined as $0 < \beta << 1$. The noise-reduced signal is calculated by the inverse fourier transform using the power spectrum of the noise-reduced signal and the phase of observed signal. However, there is a problem that the SS generates a distortion and a residual noise called musical tone in the noise-reduced signal.

2.2 Flooring processing-improved Spectral Subtraction (F-SS)

The F-SS (5) has been proposed as the effective method to reduce the musical tone. The musical tone is generated when the power difference of the remnant noise is generated in the noise-reduced signal by switching subtraction processing and flooring processing in the SS. In the F-SS, the flooring coefficient is defined as $0 << \beta < 1$ to reduce the generation of the musical tone. The F-SS further has achieved a big amount of noise reduction by iteratively performing noise reduction. Equation (2) is the process of the F-SS.

$$\begin{aligned} &|\hat{X}_i(\omega)|^2 = \begin{cases} |\hat{X}_{i-1}(\omega)|^2 - \alpha|\hat{N}_i(\omega)|^2, & \text{if } |\hat{X}_{i-1}(\omega)|^2 - \alpha|\hat{N}_i(\omega)|^2 > \beta|\hat{X}_{i-1}(\omega)|^2, \\ \beta|\hat{X}_{i-1}(\omega)|^2, & \text{otherwise,} \end{cases} \\ &|\hat{X}_0(\omega)|^2 = |Y(\omega)|^2, \qquad i = 1, 2, \cdots, I, \end{aligned} \tag{2}$$

where $|\hat{X}_i(\omega)|^2$ and $|\hat{N}_i(\omega)|^2$ indicate the power spectrum of the noise-reduced signal and that of the estimated noise at the i-th iteration, respectively, I indicates number of iteration. It is thus possible to reduce the musical tone, but it is still difficult to reduce the distortion of the noise-reduced signal.

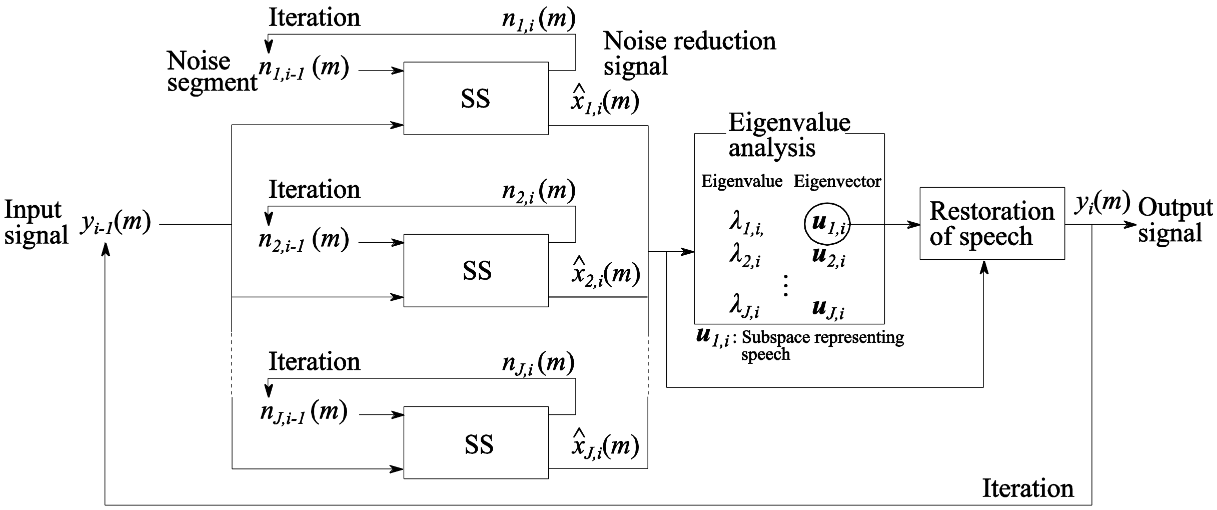
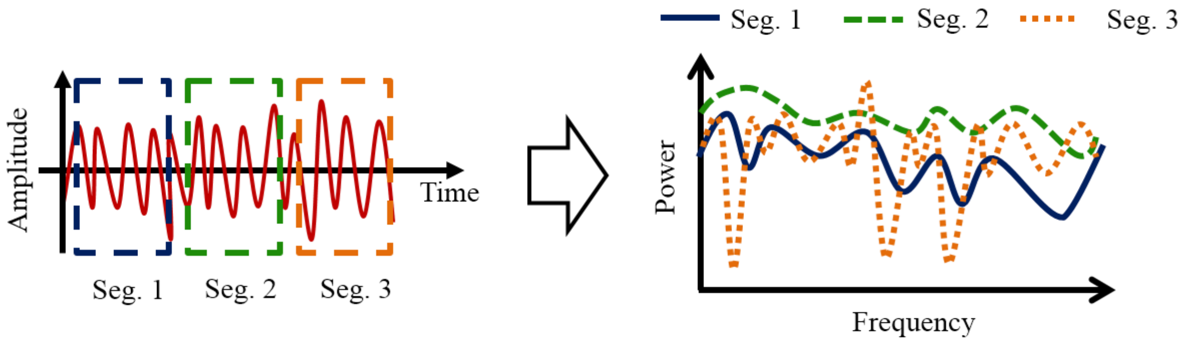


Figure 1 –The block diagram of the proposed S-SS



(a) Noise signal

(b) Estimated noise of each segment

Figure 2 – Various estimated noises in the SS

3. THE PROPOSED METHOD

3.1 Subspace-based Spectral Subtraction (S-SS)

In this paper, we propose the S-SS that reduces the distortion of the noise-reduced signal utilizing the eigenvalue analysis.

Figure 1 shows the block diagram of the proposed S-SS. In Fig. 1, $y_i(m)$ indicates input signal at the i-th iteration, $n_{j,i}(m)$ indicates noise segment at the i-th iteration, $\hat{x}_{j,i}(m)$ indicates the noise-reduced signal at the i-th iteration by the SS, $\lambda_{j,i}$ indicates the j-th eigenvalue at the i-th iteration, $\boldsymbol{u}_{j,i}$ indicates the j-th eigenvector at the i-th iteration, m indicates sample number, j indicates number of the noise-reduced signal. In the same noisy environment, different estimated noises are calculated by changing the average number of frames and noise segment as shown Fig. 2. If the SS uses these different estimated noises to reduce the unwanted noise, different noise-reduced signals are acquired. The proposed S-SS acquires multiple noise-reduced signals by the SS under the various conditions of noise estimation. These noise-reduced signals include a common speech component, a different remnant noise component and a different distortion component. To reduce the distortion, we consider that extracting the common speech component is effective. The proposed S-SS performs the eigenvalue analysis with multiple noise-reduced signals in order to calculate the subspace of the common speech component. To calculate the subspace of the common speech component, the proposed S-SS only uses the first eigenvalue and eigenvector. Eigenvalues and eigenvectors are calculated from the variance-covariance matrix as shown in Eqs. (3) and (4).

Table 1 – Experimental conditions

Speech corpus	ATR phoneme balanced sentences (7)
Speakers	Two female and two male speakers
Sampling	16 kHz, 16 bit
Window function	Hanning window
Frame length	64 ms (1024 samples)
Shift length	16 ms (256 samples)
Coefficients of F-SS	α:2.0, β:0.9
Coefficients of S-SS	α:2.0, β:0.9
Average of noise estimation in F-SS	16 frames
Average of noise estimation in S-SS	1, 2, 4, 8 and 16 frames
Noise intervals	5 intervals
SNR	0, 10 and 20 dB
Noise sources	Fan noise and factory noise (8)

$$\boldsymbol{S}_i = \begin{bmatrix} s_{1,1} & \cdots & s_{J,1} \\ \vdots & \ddots & \vdots \\ s_{1,J} & \cdots & s_{J,J} \end{bmatrix}, \tag{3}$$

$$s_{j,j} = \frac{1}{m}\sum_{m=0}^{M-1}\left(\hat{x}_{j,i}(m) - \bar{\hat{x}}_{j,i}\right)\left(\hat{x}_{j,i}(m) - \bar{\hat{x}}_{j,i}\right), \tag{4}$$

where $\boldsymbol{S}_i$ indicates the variance-covariance matrix and M indicates the total number of samples. In the proposed S-SS, the eigenvalue analysis is performed utilizing Jacobi method (6). In the proposed S-SS at the i-th iteration, the eigenvalue matrix and the eigenvector matrix are calculated by the Jacobi method as follows.

$$\boldsymbol{\Lambda}_i = \boldsymbol{U}_i^{-1}\boldsymbol{S}_i\boldsymbol{U}_i = \mathrm{diag}\left[\lambda_{1,i},\cdots,\lambda_{J,i}\right], \tag{5}$$

$$\boldsymbol{U}_i = \left[\boldsymbol{u}_{1,i},\cdots,\boldsymbol{u}_{J,i}\right], \tag{6}$$

$$\boldsymbol{u}_{j,i} = \left[u_{j,i}(1),\cdots,u_{j,i}(j),\cdots,u_{j,i}(J)\right], \tag{7}$$

where $\boldsymbol{\Lambda}_i$ indicates the eigenvalue matrix at the i-th iteration, $\boldsymbol{U}_i$ indicates the eigenvector matrix at the i-th iteration, $u_{j,i}(j)$ indicates the element of the j-th eigenvector at the i-th iteration, and J indicates the total number of eigenvalues. In multiple noise-reduced signals, the common speech component is represented by the eigenvalue with the highest contribution rate is utilized. The S-SS finally calculates the output signal which is reduced the distortion by using the first eigenvector, as shown in Eq. (8).

$$y_i(m) = \sum_{j=1}^{J} u_{1,i}(j)\hat{x}_{j,i}(m). \tag{8}$$

where $y_i(m)$ indicates the output signal by the S-SS at the i-th iteration and $u_{1,j}(j)$ indicates the element of the first eigenvector at the i-th iteration. The S-SS furthermore can reduce the unwanted noise and the distortion of the observed signal by iterating these processes.

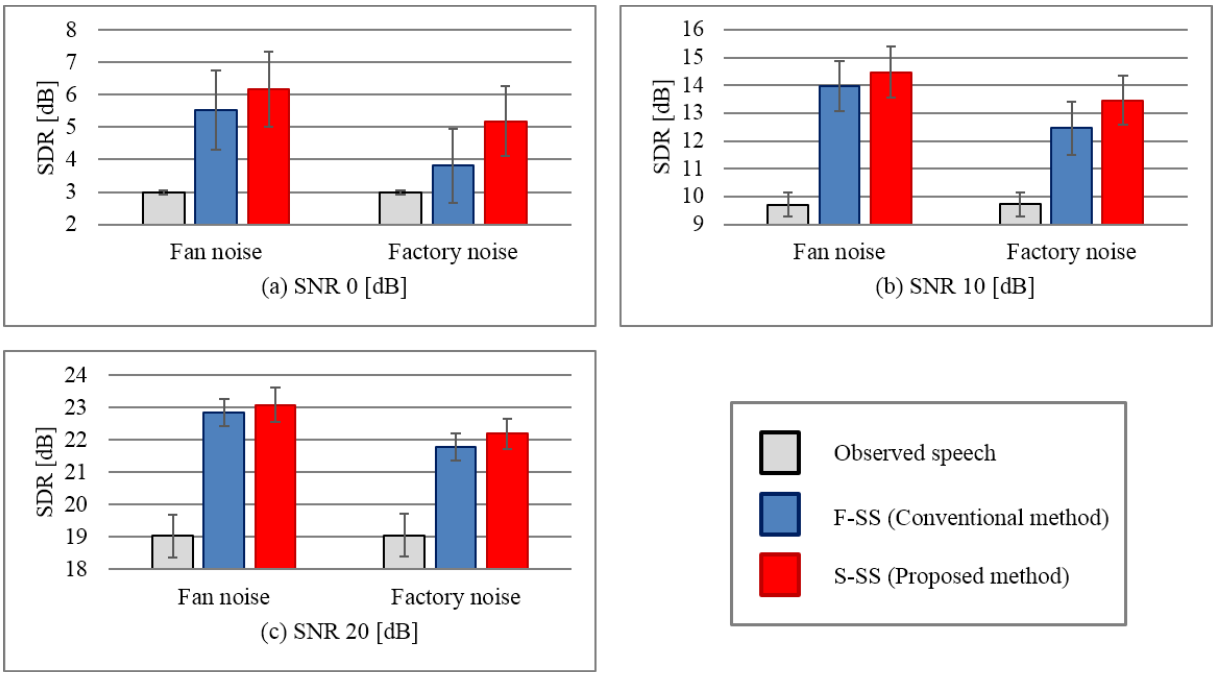


Figure 3 –Experimental results for SDR

4. EXPERIMENT

4.1 Experimental condition

The objective experiment was carried out for verifying the effectiveness of the proposed S-SS. In the objective experiment, we evaluated the amount of the distortion in the output signal by the conventional F-SS and the proposed S-SS.

Experimental conditions are shown in Tb 1. We evaluated the effectiveness of the F-SS and the proposed S-SS for fan and factory noises. These are real environmental noises. In particular, fan noise is stationary noise, factory noise is the non-stationary noise. These noises were added to the ATR phoneme balanced sentences with three kinds of SNR conditions (0, 10 and 20 dB). To set the different estimated noise used the noise reduction, the proposed S-SS used multiple noise segments and the average number of frames of noise estimation. In this experiment, we reduced the unwanted noise until the SNR of after noise reduction is 30 dB in different SNR environments.

4.1.1 Evaluation index for speech distortion

The speech distortion is evaluated based on signal-to-distortion ratio (SDR). The SDR is derived from Eq. (9).

$$\mathrm{SDR} = 10\log_{10}\left(\sum_{m=0}^{M-1} x^2(m) \Big/ \sum_{m=0}^{M-1} (x(m) - \gamma\hat{x}(m))^2\right), \quad \gamma = \sum_{m=0}^{M-1} |\, x(m)\,| \Big/ \sum_{m=0}^{M-1} |\, \hat{x}(m)\,|, \tag{9}$$

where $x(m)$ indicates the clean speech signal and $\hat{x}(m)$ indicates the noise-reduced signal. Higher SDR value indicates the lower amount of the distortion.

4.2 Experimental results

Figure 3 shows the experimental results for SDR. In Fig. 3, the horizontal axis represents the SDR value and the vertical axis represents the kind of noises (Fan noise and factory noise). Figures 3 (a) ~ (c) furthermore represent the results of each SNR (0, 10 and 20 dB), the color of the figure shows the kind of noise reduction methods (the proposed S-SS, the F-SS and the observed signal). As a result of Fig. 3, the proposed S-SS achieved higher SDR than the F-SS in fan noise and factory noise. From these results, we confirmed that the proposed S-SS can reduce the distortion of the noise-reduced signal. However, the improvement of SDR in fan noise is a small. It is necessary to improve the performance of the proposed S-SS in a various noisy environments

5. CONCLUSION

We proposed the S-SS as a new noise reduction method. The proposed S-SS reduce the unwanted noise and the distortion of the observed signal by utilizing the eigenvalue analysis. The objective experiment with SDR was carried out for evaluating the distortion and noise reduction performance of the F-SS and S-SS. As a result of objective experiment, we confirmed that the proposed S-SS can reduce the distortion of the noise-reduced signal as well as the unwanted noise.

In order to reduce the distortion of the noise-reduced signal, we should improve the eigenvalue analysis used the proposed S-SS as future work. In addition, we consider again for noise estimation to effectively reduce the non-stationary noise.

ACKNOWLEDGEMENTS

This work was partly supported by Grand-in-Aid for Scientific Research, funded by MEXT, Japan.

REFERENCES

1. J.L. Flanagan, J.D. Johnston, R. Zahn and G.W. Elko, ``Computer-steered microphone arrays for sound transduction in large rooms," The Journal of the Acoustical Society of America, Vol. 78, No. 5, pp. 1508-1518, 1985.
2. Y. Takahashi, T. Takatani, H. Saruwatari and K. Shikano, ``Blind spatial subtraction array with independent component analysis for hands-free speech recognition," Proc. InternationalWorkshop for Acoustic Echo and Noise Control 2006, CD-ROM, 2006.
3. S.F. Boll, ``Suppression of acoustic noise in speech using spectral subtraction," IEEE Transactions on Acoustic, Speech and Signal Processing, Vol. ASSP-27, No. 2, pp. 113-120, 1979.
4. S.V. Vaseghi, ``Advanced Digital Signal Processing and Noise Reduction, " John Wiley & Son Ltd, 1995.
5. T. Fukumori, M. Morise, T. Nishiura, Y. Yamashita and H. Nanjo, ``The estimation of optimum subtraction parameters for iterative spectral subtraction towards musical tone reduction," Proc. Internoise2011, PaperID: Mon-P-21, 2011.
6. W.H. Press, B.P. Flannery, S.A. Teukolsky and W.T. Vetterling, ``Numerical Recipes in C," CAMBRIDGE UNIVERSITY PRESS, 2004.
7. Y. Sagisaka, K. Takeda, M. Abe, S. Katagiri, T. Ueda, and H. Kuwabara, "A large-scale Japanese speech database, " Proc. ICSLP90, Vol. 2, pp. 1089-1092, 1990.
8. JEIDA Noise Database (JEIDA-NOISE), http://www.sunrisemusic.co.jp/database/fl/noisedata01_fl.html