Title	小規模マイクロホンアレーを用いた複数雑音にロバストな雑音除去法に関する研究
Author(s)	籠，貴司
Citation	
Issue Date	2001-09
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1549
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

A study on a noise reduction method for multi-noises using a small-scale microphone array

Takashi Kago

School of School of Information Science,
Japan Advanced Institute of Science and Technology

August 15, 2001

Keywords: small-scale microphone array, subtractive-type beamformer, multi-noises, algorithm dividing frequency band.

1 Intoroduction

There is a great demand for noise reduction tecniques with the advance of digital acoustic processing. In recent years, many noise reduction methods using a microphone array are proposed. Microphone array can compose beamformers. However, even if the method is useful in computer simulations, noise reduction accuracy decreases in real environments. It is caused by the following conditions.

(1)sudden noise

(2)multi-noises coming from different directions

(3)reverberation

For the problem(1), Mizumachi proposed a method using a 3ch. microphone array[1]. This method is robust to sudden noise by subtracting estimated noise from a received signal frame by frame. However, since the method assumes the number of noises is one, the method can not reduce multi-noises.

In this study, we propose a new noise reduction method being robust to multi-noises by dividing frequency band and using Mizumachi's proposed method in each frequency band. The method is robust to both sudden noise and multi-noises and adaptable in real environments. In addition, a small-scale microphone array is used in the method, supposing a front-end of hands-free Automatic Speech Recognizer (ASR).

2 Noise reduction algorithm

In this chapter, we give an outline of our proposed method. The basic algorithm is the same as that of Mizumachi's proposed method. A characteristic of our proposed method is dividing frequency band and estimating the largest noise spectrum in each band.

2.1 Estimation of arrival time delay of noises

Assuming that the target signal comes from the front, and $N_k(\tilde{\omega})$ is a noise coming from a direction such that the arrival time delay is δ_k between neighboring microphones, which is the largest in the frequency band $\omega_{k-1} < |\tilde{\omega}| \leq \omega_{k+1}$ ($k = 1 \sim K$). Using the signal $l(t)$ (in the left microphone) and $c(t)$ (in the center microphone), and the $c(t)$ and $r(t)$ (in the right microphone), signals $G_{lc}(\tilde{\omega})$ and $G_{cr}(\tilde{\omega})$ reduced the target signal (subtractive-type beamformer) are calculated.

$$G_{xy}(\tilde{\omega}) = \mathrm{FFT}\left[\frac{\{x(t+\tau) - x(t-\tau)\} - \{y(t+\tau) - y(t-\tau)\}}{4}\right] \qquad \omega_{k-1} < |\tilde{\omega}| \leq \omega_{k+1} \quad (1)$$

Here the combination of xy is lc, cr or lr. τ is a certain constant except 0, which can control a focus. The focus is a direction abstracting the target signal. The arrival time delay δ_k is estimated by crosscorrelation of $G_{lc}(\tilde{\omega})$ and $G_{cr}(\tilde{\omega})$.

2.2 Estimation of noise spectrum

The noise spectrum $N_k(\tilde{\omega})$ in the center microphone is estimated using δ_k, $G_{lc}(\tilde{\omega})$ and $G_{cr}(\tilde{\omega})$. By substituting δ_k into τ in $G_{lr}(\tilde{\omega})$, and $\frac{\delta_k}{2}$ into τ in $G_{cr}(\tilde{\omega})$, the focus is taken the noise direction.

$$G_{lr}(\tilde{\omega}) = N_k(\tilde{\omega}) \sin^2 \tilde{\omega}\delta_k \qquad \omega_{k-1} < |\tilde{\omega}| \leq \omega_k \quad (2)$$

$$G_{cr}(\tilde{\omega}) = N_k(\tilde{\omega}) e^{j\tilde{\omega}\frac{\delta_k}{2}} \sin^2 \tilde{\omega}\frac{\delta_k}{2} \qquad \omega_{k-1} < |\tilde{\omega}| \leq \omega_k \quad (3)$$

In the case such that $\tilde{\omega} = \frac{n\pi}{\delta_k}$ (n:integer) in $G_{lr}(\tilde{\omega})$ and n is an even number in $G_{cr}(\tilde{\omega})$, the noise spectrum can not be estimated accurately. Then, two parameters ε_1 and ε_2 are assumed. The noise spectrum $\hat{N}_k(\tilde{\omega})$ can be estimated by Eq.(4).

$$\hat{N}_k(\tilde{\omega}) = \begin{cases} G_{lr}(\tilde{\omega}) / \sin^2 \tilde{\omega}\delta_k & \sin^2 \tilde{\omega}\delta_k > \varepsilon_1 \\ G_{cr}(\tilde{\omega}) e^{-j\tilde{\omega}\frac{\delta_k}{2}} / \sin^2 \tilde{\omega}\frac{\delta_k}{2} & (\sin^2 \tilde{\omega}\delta_k \leq \varepsilon_1) \wedge \left(\sin^2 \tilde{\omega}\frac{\delta_k}{2} > \varepsilon_2\right) \\ G_{lr}(\tilde{\omega}) & \sin^2 \tilde{\omega}\frac{\delta_k}{2} \leq \varepsilon_2 \end{cases} \quad (4)$$

By applying the above estimations in each divided frequency band, we can get the largest noise $\hat{N}_1(\tilde{\omega}) \sim \hat{N}_K(\tilde{\omega})$. $\hat{N}(\omega)$ is the noise spectrum combined them in all frequency band.

2.3 Reduction of noise spectrum

Finally, the noises are reduced by subtracting amplitude spectrum of the estimated noise $|\hat{N}(\omega)|$ from that of the signal in the center microphone $|C(\omega)|$ using non-liner Spectral Subtraction (SS). SS is proposed by Boll[3], and non-liner SS is a improved method on SS.

$$|\hat{S}(\omega)| = \begin{cases} |C(\omega)| - \alpha|\hat{N}(\omega)| & |\mathrm{C}(\omega)| \geq \alpha|\hat{\mathrm{N}}(\omega)| \\ \beta|C(\omega)| & \text{otherwise} \end{cases} \quad (5)$$

α and β the subtraction coefficient and the flooring coefficient.

3 Noise reduction experiment

To evalute the noise reduction accuracy of our proposed method, experiments using multi-noises are performed. 3ch. equally-spaced liner microphone array, that the neighboring microphones space is 10 cm, is used for the experiments.

3.1 Experimental conditions

The experiments are performed in a sound proof room. The reverbration time is about 50 msec over 500 Hz, and 100～300 msec in lower frequency band than 500 Hz. The speaker presenting the target signal is located in the front, 0.5 m from the microphone array. The speaker presenting the noise1 and the noise2 are located 45 °to the left and 30 °to the right respectively, and the speakers for the noises are 2.0 m from the microphone array. The target signal is a japanese word /aoi/ uttered by a male speaker 'mau' in the ATR speech database. The noise1 is a sound of gas spray (band noise in 4～8 kHz), and noise2 is a beep of a cellular phone (frequency ranges having large powers are 2.8 and 5.6 kHz) in the RWCP sound scene database. These sounds are presented simultaneously in the room. The sound data are sampled at 48 kHz with 16 bit accuracy. The parameter values for the experiments are shown in Table.1. Since the frequency band is divided as fine as possible, the noise direction and spectrum estimation are achieved in 48000/1024 Hz steps.

Table 1: parameter value

parameter	value
frame length	1024 pt (21.3 msec)
frame period	512 pt (10.6 msec)
window type	Hanning window
subtraction coefficient α	1.0
flooring coefficient β	0.001
ε_1	0.5
ε_2	0.2

For the evalution, the following SNR and LSD are used. LSD is calculated only in the frame existing the speech signal. $s(t_n)$ and $S(\omega)$ are an original sound and $\hat{s}(t_n)$ and $\hat{S}(\omega)$ are a noise-reduced sound. The unit is both dB, and $W = 6$ kHz.

$$\mathrm{SNR} = 10\log_{10}\frac{\sum_n s^2(t_n)}{\sum_n \{s(t_n) - \hat{s}(t_n)\}^2} \tag{6}$$

$$\mathrm{LSD} = \sqrt{\frac{1}{W}\sum_{\omega}^{W}\left(20\log_{10}\frac{|S(\omega)|}{|\hat{S}(\omega)|}\right)^2} \tag{7}$$

3.2 Experimental results

The experimental results are shown in Fig.1. The figure shows (a)noise-free speech, (b)noise-added speech (c)estimated noise by our proposed method, (d)noise-reduced speech by our proposed method, and (e)noise-reduced speech by Mizumachi's proposed method. Here, noise-free speech is a sound recorded by the center microphone when only the speech is presented.

The figure shows that our proposed method can estimate the two noises spectrums clearly(Fig.1(c)), and can reduce almost the noises(Fig.1(d)). However, the noise of a cellular phone in 5.6 kHz can not be reduced by Mizumachi's proposed method.

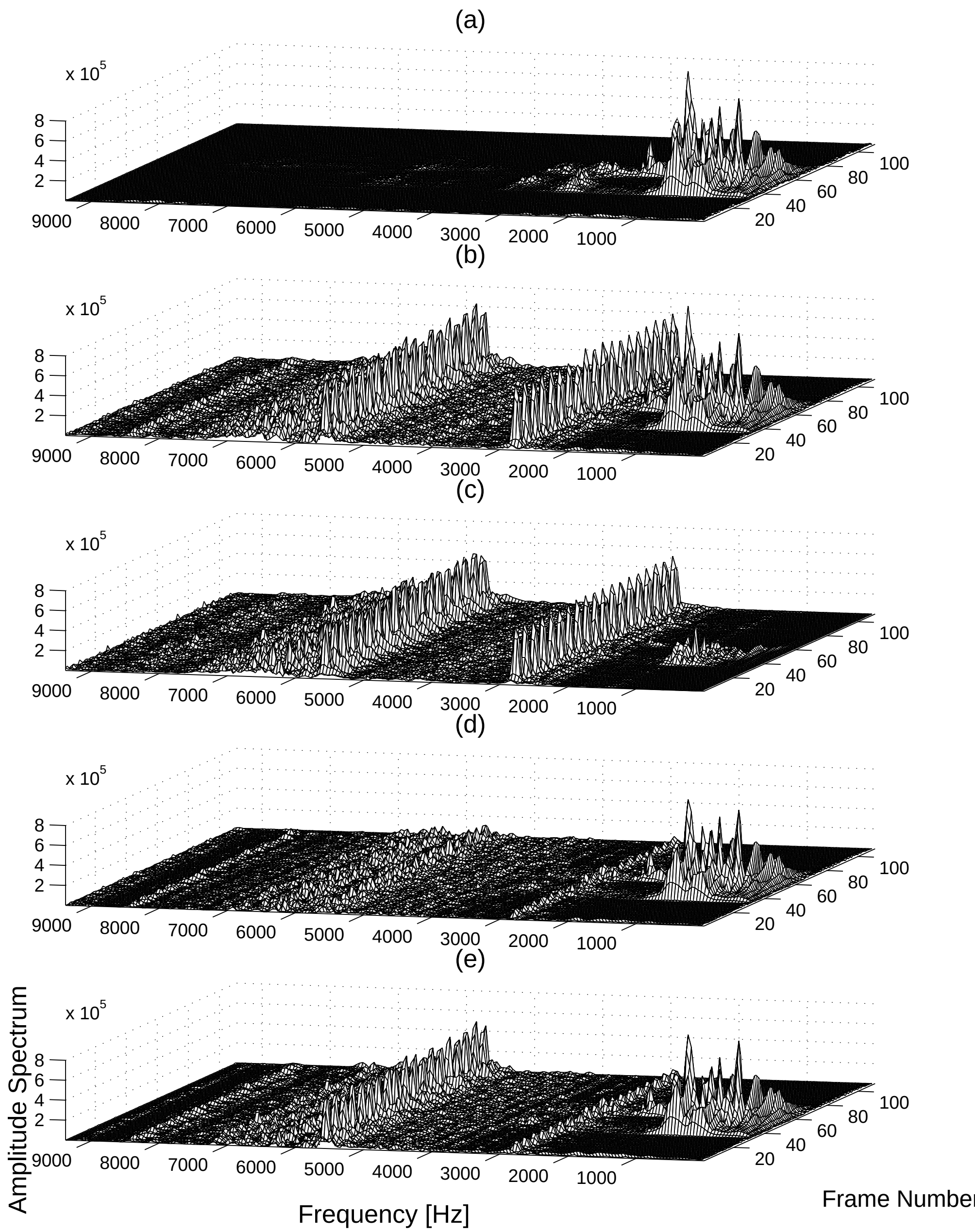

Figure 1: experiment results —spectrum— (a)noise-free speech (b)noise-added speech (c)estimated noise by our proposed method (d)noise-reduced speech by our proposed method (e)noise-reduced speech by Mizumachi's proposed method

Table.2 shows the values of SNR and LSD before and after noise reduction. The SNR is increased about 10 dB and the LSD is decreased 5 dB. This indecates that our proposed method seems to be robust to multi-noises. In addition, superiority of our proposed method is evident by a comparizon of our method with Mizumachi's proposed one.

Table 2: evalution by SNR and LSD

	before noise reduction	our proposed method	Mizumachi's proposed method
SNR [dB]	-4.0	6.5	0.9
LSD [dB]	17.9	12.8	14.4

4 Conclusion

We intoroduced the algorithm dividing frequency band into Mizumachi's proposed method, and proposed a new noise reduction method being robust to multi-noises. The noise reduction experiment showed that our proposed method is useful for multi-noises and superior to Mizumachi's proposed method.

References

[1] Mitsunori Mizumachi, "A study on a noise reduction method by paired microphone", JAIST doctor thesis (2000)

[2] Masato Akagi and Mitsunori Mizumachi, "NOISE REDUCTION BY PAIRED MICROPHONES", Proc.EUROSPEECH '97, Rhodes, Greece, Vol.1, pp.335-338 (1997)

[3] S.F.Boll, "Suppression of Acoustic Noise in Speech Using Spectral Subtraction", IEEE Trans.ASSP, Vol.27, No.2, pp.113-120 (1979)