Title	雑音抑圧処理を付加した音源方向推定に関する研究
Author(s)	小林，利秋
Citation	
Issue Date	2001-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1472
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

A study on the estimation of sound source direction equipped with noise reduction process

Toshiaki Kobayashi

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

February 15, 2001

Keywords: sound source direction estimation, microphone-array, noise reduction, reverberation, dynamic threshold.

1 Introduction

Estimations of sound source direction have a wide range of application today in various fields, for example, hands-free speech recognition, teleconferencing system. Conventional estimation methods of source direction exhibit performance degradations in noisy and reverberant environment[1]. Therefore, this paper proposes a new estimation method of sound source direction. The proposed method is equipped with noise reduction process by using harmonic structure of the target speech. The noise reduction process is combined with Nishida's direction estimation method that is robust in multipath environment[2]. The combination can be expected to be robust system in noisy and reverberant environments.

2 Microphone array

The proposed estimation method of sound source direction is performed by using a microphone array that arranges microphones in the apexes of the equilateral triangle. This is the smallest number of microphones that can estimate all direction in the two-dimensional plane. This microphone-array can solve the problem of front-back miss judgment when using a linear microphone-array.

3 Outline of the proposed method

An overview of the proposal method is shown in the following.

- noise reduction process
  - fundamental frequency (period) estimation
  - time-frequency filtering
- onset detection by using dynamic threshold
  - peek extract
  - onset detection
- detection of time difference
- converting time difference into direction
- eliminating anomalies
  - direction integration
  - smoothing in time domain

4 Noise reduction process

The first, fundamental period of the target signal is estimated by the Noll's Cepstrum method from the signal received by the microphones. Then, a pitch-based time-frequency filtering[3] is applied to the noise reduction process.

4.1 Formulation

Assume that the target signal $s(t)$ is a harmonic complex tone and $n(t)$ is noise. Thus, the signal received by the microphone is

$$x(t) = s(t) + n(t) \quad (1)$$
$$= \sum_n a_n e^{j(n\omega_0 t+\theta_n)} + \sum_k b_k e^{j(n\omega_k t+\theta_k)} \quad (2)$$

where T is the fundamental period. If T is constant, a signal $g(t)$, which is the subtracted signal shifted $x(t)$ to $\pm T$ in time from $x(t)$, is represented as follows.

$$g(t) = \frac{2x(t) - x(t-T) - x(t+T)}{4} \quad (3)$$
$$= \sum_k b_k e^{j(n\omega_k t+\theta_k)} \sin^2 \frac{\omega_k}{\omega_0}\pi \quad (4)$$

Fourier transform $G(\omega_k)$ of $g(t)$ is

$$G(\omega_k) = N(\omega_k)\sin^2\frac{\omega_k}{\omega_0}\pi, \tag{5}$$

where $N(\omega_k)$ is Fourier transform of $n(t)$. Then, the nosie spectrum $N(\omega_k)$ is represented as

$$N(\omega_k) = G(\omega_k)/\sin^2\frac{\omega_k}{\omega_0}\pi \tag{6}$$

If ω_k/ω_0 is an integer, the noise spectrum $N(\omega_k)$ in equation (6) is infinity. For practical applications, a parameter ε is established.

$$N(\omega_k) = \begin{cases} G(\omega_k)/\sin^2\frac{\omega_k}{\omega_0}\pi & \text{for } |\sin\frac{\omega_k}{\omega_0}\pi| \geq \varepsilon \\ 0 & \text{for } |\sin\frac{\omega_k}{\omega_0}\pi| < \varepsilon \end{cases} \tag{7}$$

The target signal $s(t)$ can be estimated by subtracting the noise signal $n(t)$, which is inverse Fourier transform of $N(\omega_k)$, from $x(t)$. This means a comb filter that can control pass bands by a value of ε.

Fig.1 shows a block diagram of the above process.

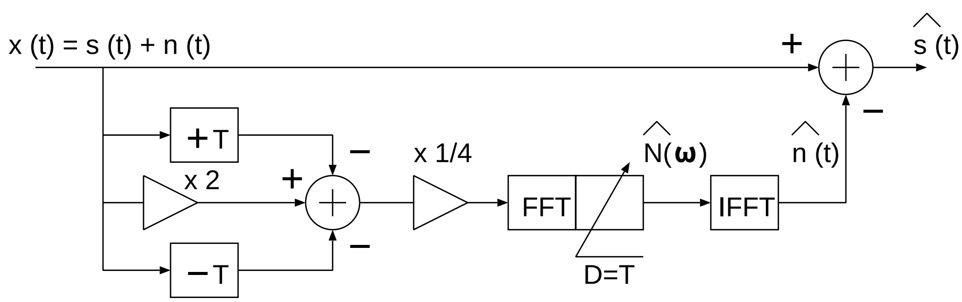

Figure 1: Block diagram of time - frequency filtering

4.2 Nishida's onset detector by using dynamic threshold

The onset of the noise suppression signal is detected as follow. The detector proposed by Nishisa is is applied to the onset detector. The dynamic threshold as shown the below is proposed. The dynamic threshold detects onset of the signal with changing threshold level. That is:

- An onset is detected when amplitude sets over a threshold
- The amplitude at that time is hold as an initial value of a threshold level.
- The threshold level decreases exponentially with time.
- Wherever amplitude gets over a threshold, the initial value of the threshold reset.

The above behaviors are shown in Fig. 2 for example. The dynamic threshold does not detect reflected sound, which delays from the direct sound and the power is smaller than the direct sound.

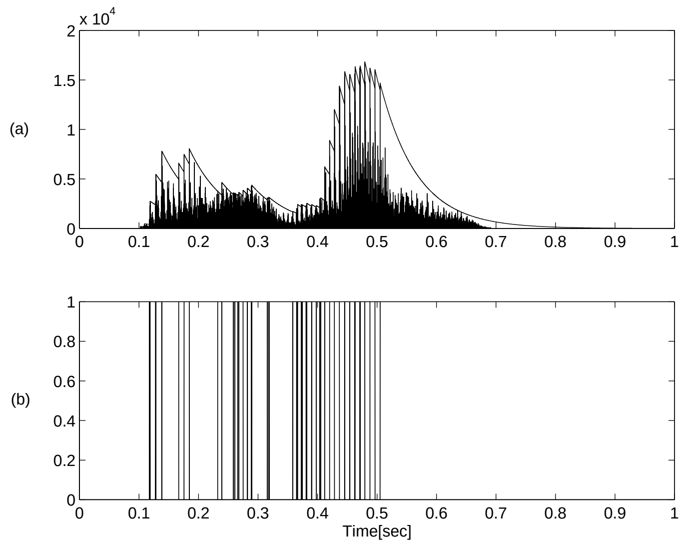

Figure 2: dynamic threshold & onset detect

5 Simulations and results

Simulations to confirm the ability of the proposed method were carried out under various conditions. The comparative experiment with the Crosspower Spectrum Phase (CPSP) method[1] that is used for usual research, and Nishida's method is performed in the simulations.

The results show that the proposed method outperforms the other two in the presence of arriving noise, which is shown in Fig. 3. The simulation parameters are shown by Table 1. % anomalies in Fig. 3 is represented that the number of detection points outside five degrees to the sound source direction is divided by the number of all detected points. This simulation result shows superior effects of the noise reduction pre-processing.

Table 1: simulation parameters

target signal	word speech of ATR database, 18 samples arriving angle : setting at random
noise	white noise arriving angle : 60 ° SNR: 10dB～30dB
1st echo	echo-to-signal ratio : -4dB arrival delay: 2ms
2nd echo	echo-to-signal ratio : -8dB arrival delay: 10～15ms

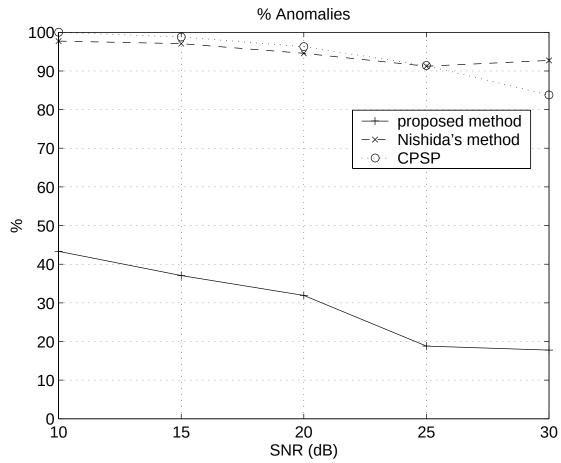

Figure 3: simulated result in the presence of arriving noise & echoes

The other experiment for simulating room acoustics[4] is performed. Assume the room conditions are illustrated by Fig. 4, and the results are shown in Fig. 5.

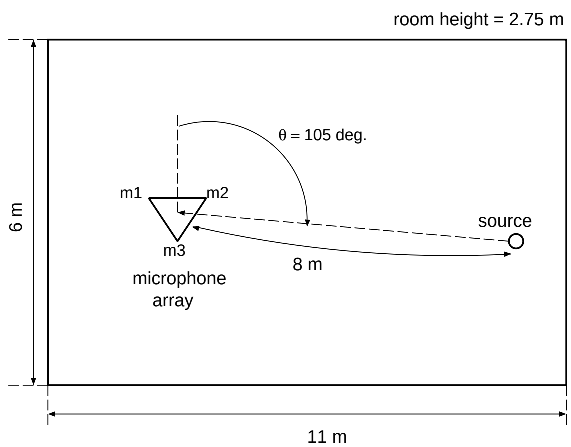

Figure 4: simulated room

The results show the proposed method outperform the other two methods in short reverberation time, but the performance of the proposed method reduced when reverberation time is longer than 200ms.

6 Conclusion

The estimation method of sound source direction equipped with noise reduction process has been proposed. The proposed method excels in the presence of arriving noise, which is confirmed as results of the comparison with the conventional method.

The other results of the experiment using simulating room acoustics show the performance deteriorated when reverberation time is longer than 200ms. This problem will be

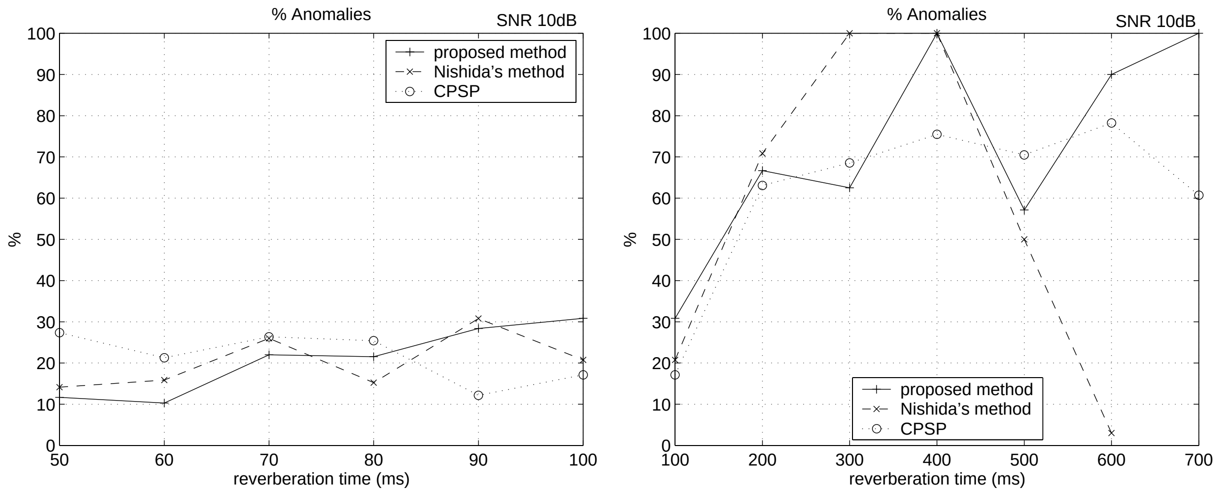

Figure 5: experimental results for simulating room acoustics : SNR 10dB

a next study hereafter.

References

[1] Maurizio Omologo, "Acoustic Event Localization Using A Crosspower-Spectrum Phase Based Technique", Proceedings of ICASSP 94, vol.2, pp. 273-276, Adelaide, 1994.

[2] T.Nishida ,"Research on the estimation of sound source direction using multi microphone," JAIST master thesis, 1999.

[3] M.Akagi et.al., "Speech Enhancement and Segregation based on Human Auditory Mechanisms," Proc, Int. Conf. Information Society in the 21st Century(IS2000), pp.102-126, Aizu-Wakamatsu, November 2000.

[4] Jont B. Allen, David A. Berkley, "Image Method for Efficiently Simulating Small-Room Acoustics," J.Acoust.Soc.Am. 65(4), pp. 954-950, April 1979.