Title	多点受音信号を用いた残響除去法に関する基礎的研究
Author(s)	水町，光徳
Citation	
Issue Date	1997-03
Type	Thesis or Dissertation
Text version	author
URL	http://hdl.handle.net/10119/1039
Rights	
Description	Supervisor:赤木 正人，情報科学研究科，修士

A Fundamental Study of Dereverberation Using Multi-microphone

Mitsunori MIZUMACHI

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

February 14, 1997

Keywords: Dereverberation, Microphone Array, Spectral Subtraction.

1 Introduction

A large number of current speech recognition systems for clean speech display great abilities in ideal environments which have no noises and reverberations. However, there are many literatures, e.g. [1] reported that performances of these speech recognition systems go down extremely in real environments because noises and reverberations distort speech.

Therefore, this thesis proposes a new method of dereverberation as a front-end posessor of speech recognition systems to support making speech recognition systems practicable by reducing the distortions of amplitude spectrums caused by the reverberations.

2 Algorithm Description

A new dereverberation method proposed here uses a microphone array constructed with three linear-equally-spaced microphones as an input device.

This method detects revereberant ingredients analytically at the position of the center microphone by using two signals received by the left and right microphones of the microphone array. Then, dereverberation can be accomplished by subtracting revereberant ingredients from the signal received by the center microphone which includes both a direct wave and its revereberant ingredients.

2.1 Detection of the Reverberant Ingredient

This method adopts a sound source segregation model[2] proposed by Imada et al. to detect the reverberant ingredient. The source segregation model can detect an objective signal precisely coming from the direction that the signal arrival time between two microphones is different, and reduce a signal perfectly coming from the front.

Hence, the directivity of the microphone array constructed by the sound source segregation model were investigated. The result shows that signals coming from all directions except the front can be detected roughly in low and mid frequency range below 1 kHz using this microphone array.

Thus, in this thesis, a signal detection algorithm to detect the reverberant ingredient based on sound source segregation model is proposed, assuming that the objective signal comes from the front. The following is a concept of the signal detection algorithm to detect the reverberant ingredient.

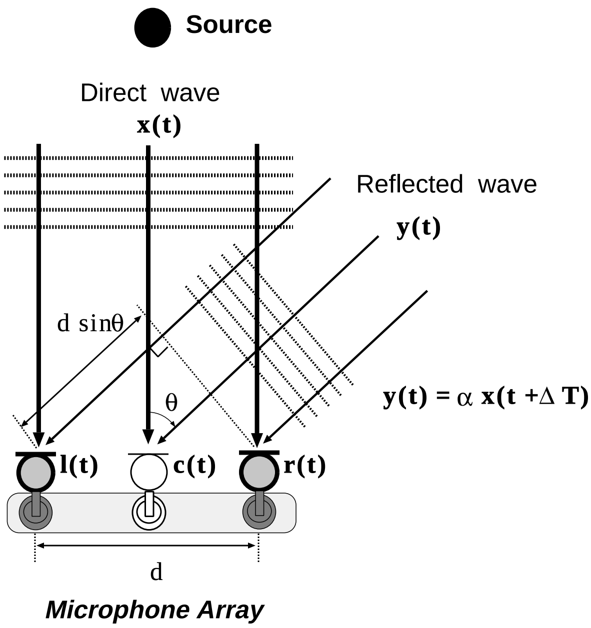

Figure 1: Illustration to explain a signal detection algorithm

Let us assume that a direct wave $x(t)$ comes from the front and one of reflected waves $y(t)$ comes from the direction such that the difference of arrival time between two microphones locating at both ends of a 3ch. microphone array is 2δ as shown in Figure 1. The arrival sound at each microphone is assumed as follows,

$$\left\{ \begin{array}{lcl} l(t) & = & x(t) + y(t-\delta) \\ c(t) & = & x(t) + y(t) \\ r(t) & = & x(t) + y(t+\delta), \end{array} \right. \tag{1}$$

where $l(t)$, $c(t)$ and $r(t)$ are signals received by left, center and right microphones of the microphone array.

The relationship between an objective signal $y(t)$ which is the reverberant ingredient and the received signals $l(t)$ and $r(t)$ can be formulated not using the direct wave $x(t)$ but using the following equation;

$$g(t) = \{l(t+\tau) - l(t-\tau)\} - \{r(t+\tau) - r(t-\tau)\}. \tag{2}$$

The signal $g(t)$ is transformed to $G(\omega)$ by the short term Fourier transformation (STFT). Then, $G(\omega)$ is calculated by using the relationship of Equation (1),

$$G(\omega) = 4Y(\omega)\ \sin\omega\delta \sin\omega\tau, \tag{3}$$

where $Y(\omega)$ is the STFT of $y(t)$, and τ is the focus of the microphone array. If the direction from which the objective signal comes can be known, the estimated signal $\hat{y}(t)$ of the objective signal $y(t)$ at the position of the center microphone can be calculated by adjusting the focus to the direction from which the objective signal comes. The estimated signal $\hat{y}(t)$ is

$$\hat{y}(t) = \begin{cases} \mathrm{IFFT}\left[\dfrac{G(\omega)}{4\sin^2\omega\tau}\right], & 4\sin^2\omega\tau > 0.02, \\ 0, & otherwise. \end{cases} \tag{4}$$

To decide the focus of the microphone array, this method adopts a well-known algorithm based on the time difference of signals arriving at the left and right microphones.

2.2 Removal of the Reverberant Ingredient

After detecting the reverberant ingredients, it is necessarry to subtract them from a reverberant signal received by the center microphone.

Since this method aims at processing dereverberation as a front-end processor for automatic speech recognition systems, it is enough to deal with amplitude spectrums only. For the subtraction of amplitude spectrums, this method employs the conventional spectral subtraction proposed by Boll [3].

The original spectral subtraction proposed by Boll didn't take account of subtracting the time variant signals, but this method needs to deal with speech which are the time variant signals. Therefore, this method proposes the short time frame processing and the subtraction coefficient renewal. The coefficient is renewed frame by frame using the correlation coefficient between the detected reverberant ingredient in a certain frame and the estimated direct wave in the previous frame. This technique can improve this method to cope with dereverberation of speech.

3 Experiments and Results

In this study, reverberant speech data are made by presenting some of clean speech data in the ATR speech database through a loud speaker in real reverberant room and by recording using the 3ch. microphone array.

The evaluatation measure for this method is amount of spectral distortion (SD) decrease from 100 Hz to 6 kHz frequency range. The averages of SD decrease among frames for a vowel /a/ and a word /bunri/ are shown in Table 1.

Table 1: Average SD decreases.

Data	Average SD decrease
/ a /	3.18 [dB]
/ bunri /	1.29 [dB]

The results indicate that this method works well as a front-end dereverberation processor, because Wang et al. has reported that the dereverberation is effective even SD decrease is only 1 dB[4].

4 Conclusion

This thesis proposed a method of dereverberation using a 3ch. equally-spaced linear microphone array, and its distinctive feature is the use of information about signal directions.

As the results of dereverberation experiments, this method is very useful in vowel frames. On the other hand, in consonant frames of speech, distortion of speech increases unfortunately. This problem should be solved in future.

References

[1] M. Tobita, N. Sugamura: "Effects of environmental conditions on speech recognition," J. Acoust. Soc. Jpn(J)., vol. 51, no. 4, pp. 331-335 (1995)

[2] S. Imada: "On study of sound source segregation using two microphones," JAIST Master's Thesis (1996)

[3] S. F. Boll: "Suppression of Acoustic Noise in Speech Using Spectral Subtraction," IEEE Trans. ASSP-27 (2) (1979)

[4] H Wang, F. Itakura: "Recovering of reverberated speech by multi-microphone envelope estimation method — About the influence of microphone arrangement and noise," Proc. ASJ meeting, 1-3-6 (1989.9)