Noise Reduction using Paired-microphones on Non-equally-spaced Microphone Arrangement

Mitsunori Mizumachi and Satoshi Nakamura

ATR Spoken Language Translation Research Laboratories
2-2-2 Hikaridai, "Keihanna Science City" Kyoto 619-0288 Japan
E-mail: mitsunori.mizumachi@atr.co.jp

Abstract

A wide variety of microphone arrays have been developed, and the authors have also proposed a type of equally-spaced small-scale microphone array. In this approach, a paired-microphone is selected at each frequency to design a subtractive beamformer that can estimate a noise spectrum. This paper introduces a non-equally-spaced microphone arrangement, which might give more spatial information than equally-spaced microphones, with two criteria for selecting the most suitable paired-microphone. These criteria are based on noise reduction rate and spectral smoothness, assuming that objective signals are speech. The feasibility of both the non-equally-spaced array and the criterion on spectral smoothness are confirmed by computer simulation.

1. Introduction

In recent years, speech interfaces have become more important as automatic speech recognition (ASR) is put to practical use. However, in contrast to the human auditory system, which can select only the objective signal accurately, conventional speech input devices transmit every signal as it is. Therefore, noise in the real world is an annoying problem for ASRs [1]. From various points of view, noise reduction has been aggresively investigated over the past several decades. Spectral subtraction is a typical method of noise reduction for a single input signal [2]. On the other hand, a microphone array can also be used for speech enhancement [3].

The study of microphone arrays originated in a conventional delay-and-sum beamforming approach, and the current technology has moved to a subtracting beamforming approach [4]. The subtractive beamformer would seem to offer great advantages to a small-scale microphone array. Almost all subtractive beamformers use adaptive filters to steer nulls in a beampattern toward the directions of interfering signals with no distortion toward to the direction of an objective signal. They work very well under ideal conditions, but the performance generally goes down when the number of interfering sound sources is equal to or exceeds the number of microphones. Therefore, the authors proposed another subtractive beamformer that is based on paired-microphones [5]. This method designs a beampattern without usual adaptive filters. It has been shown that this method works well even in non-stationary noise conditions with three equally-spaced microphones.

It is important for small-scale microphone arrays to make the best use of spatial information. If the distance between neighboring microphones is different, the acquired information must be different. As for the microphone arrangement, non-equally-spaced microphones might yield better performance in noise reduction than equally-spaced microphones, because it has more spatial information. This paper proposes a non-equally-spaced microphone array for noise reduction by paired-microphones. This means that there are several kinds of paired-microphones with various spacings, and then the most suitable paired-microphone must be selected properly under any circumstance. A criterion for the paired-microphone selection is also proposed and its feasibility is confirmed.

This paper is organized as follows. Section 2 describes an outline of the previously proposed noise reduction algorithm by using paired-microphones [5]. In section 3, we discuss a non-equally-spaced microphone arrangement and introduce a criterion based on spectral smoothness to select the most suitable paired-microphone. In section 4, we verify the feasibility of the proposed non-equally-spaced microphone array and the criterion. Finally, our conclusions are given in section 5.

2. Noise Reduction by Paired-microphone

The authors earlier proposed an algorithm for noise reduction using a 3-ch linear microphone array [5]. The microphone array consists of three linearly arranged microphones with 10-cm spacing. This method works by combining spatial filtering and frequency filtering and consists of three modules: direction finder, noise spectrum estimator by subtractive beamformers, and spectral subtraction.

2.1. Direction Finder

Assuming that signal directions are obtained in advance, we can design a subtractive beamformer analytically without using adaptive filters. It is a very natural scenario, but is quite difficult to make a robust direction finder in noisy environments where noise reduction is required. However, the author constructed a robust direction finder by integrating the subtractive beamformer with a traditional cross-correlation method [6].

If three microphones are prepared, the two directions corresponding to the most and the second most dominant signals can be estimated as follows. First, the direction of the most dominant signal is estimated based on the whitened cross-correlation [7]. Next, we obtain two signals that suppress the most dominant signal completely by using the subtractive beamformer with a single notch. In the case of a 3-ch microphone array, a subtractive beamformer is constructed by using two received signals by left and center microphones, and another beamformer is constructed by using two received signals by center and right microphones. The design method of the subtractive beamformer is described in the next section. The second direction is also estimated based on the whitened cross-correlation between the outputs of two beamformers.

The direction of an objective source is chosen between two estimated directions. If certain conditions for the objective source are given initially, the direction finder can track the movement of the sound source.

2.2. Subtractive Beamformer

A subtractive beamformer with a single sharp notch is designed using two signals received by a paired-microphone for eliminating a directional signal. Let us assume that there are two signals $x_1(t)$ and $x_2(t)$. The signal received by each microphone is described as follows.

$$\text{left mic.:} \quad l(t) = x_1(t-\zeta) + x_2(t-\delta), \tag{1}$$
$$\text{right mic.:} \quad r(t) = x_1(t+\zeta) + x_2(t+\delta), \tag{2}$$

where 2ζ means the arrival time difference between the left and right microphones for the signal $x_1(t)$, and 2δ for the signal $x_2(t)$. If we know that 2δ corresponds to the direction of a signal which should be eliminated, the beamformer $g_{lr}(t)$ can eliminate the signal $x_2(t)$.

$$\begin{aligned} g_{lr}(t) = \; & \tfrac{1}{4}\Big[\big\{l(t+\delta+\tau) - l(t+\delta-\tau)\big\} \\ & -\big\{r(t-\delta+\tau) - r(t-\delta-\tau)\big\}\Big], \end{aligned} \tag{3}$$

where τ is an arbitrary constant except zero.

Then the short-term Fourier transformation $G_{lr}(\omega)$ of the beamformer $g_{lr}(t)$ can be calculated as

$$G_{lr}(\omega) = X_1(\omega)\, \sin\omega(\zeta-\delta)\, \sin\omega\tau, \tag{4}$$

where $X_1(\omega)$ is the short-term Fourier transformation of the signal $x_1(t)$. Note that $G_{lr}(\omega)$ has no term that concerns the signal $x_2(t)$. Setting τ as $\zeta-\delta$, $X_1(\omega)$ can be estimated as follows.

$$X_1(\omega) = G_{lr}x(\omega)/\sin^2\omega(\zeta-\delta), \tag{5}$$
$$\omega(\zeta-\delta) \neq n\pi \quad (n\text{: integer}). \tag{6}$$

Here, assuming that $x_1(t)$ and $x_2(t)$ are an aggregate of noise signals and an objective signal, $X_1(\omega)$ gives the estimate of noise spectrum with no influence of the objective signal.

2.3. Spectral Subtraction

Noise reduction is completed by subtracting the estimated noise spectrum from the spectrum of the received signal. Therefore, spectral subtraction can be conducted without a speech period detector.

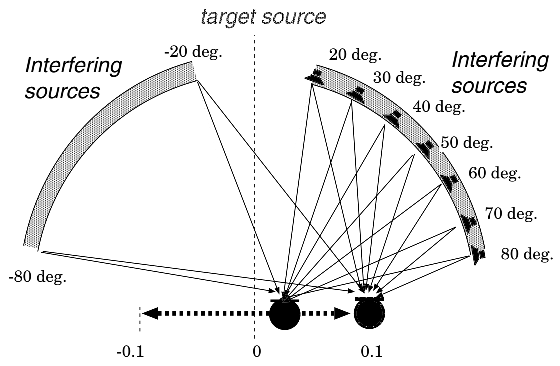

Figure 1: Experimental configuration with two microphones. One is fixed at 0.10m, and the other is moved from -0.10 m to 0.09 m at intervals of 0.01 m.

3. Non-equally-spaced Microphone Arrangement

Noise reduction can be conducted even by non-equally-spaced microphones, since a paired-microphone gives an estimate of a noise spectrum in each frequency bin in Eq. (6). However, accuracy of estimation depends on each microphone-pair.

In the case of three microphones, only two different spacings between microphone-pairs are available in an equally-spaced arrangement, but three different spacings are obtained in a non-equally-spaced arrangement. Generally speaking, wide spacing is better than narrow spacing in a low-frequency range, and the opposite applies in a high-frequency range. It is reasonable to expect that a non-equally-spaced arrangement is better than a former equally-spaced arrangement in noise reduction performance, because more suitable spacings are provided at each frequency.

3.1. Optimized Microphone Arrangement

For two received signals $x_i(t)$ and $x_j(t)$, spectrum $Y(\omega)$ of an objective signal is estimated. Noise Reduction Rate (NRR) is calculated as the Euclidean distance between a received signal $X_1(\omega)$, which includes an objective signal and noise signals, and an estimated objective signal $Y(\omega)$ in a log spectral domain.

$$NRR_\omega(x_i, x_j; y) = 20\log_{10}\frac{X_i(\omega)}{Y(\omega)} \quad [\text{dB}] \tag{7}$$

If three microphones are provided in a non-equally-spaced arrangement, three NRRs are obtained in each frequency ω. The most suitable microphone arrangement is decided as follows over various microphone arrangements.

$$\sum_{\omega} \max_{\{i,j\}=\{1,2,\cdots,N\}, i\neq j} NRR_\omega(x_i, x_j), \tag{8}$$

where max is an operator to select the maximum NRR, and N is the number of microphones.

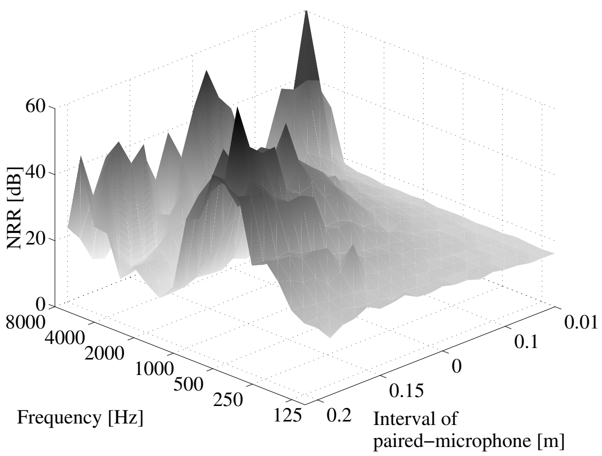

Figure 2: Noise Reduction Rate (NRR) plotted against each spacing of the paired-microphone and frequency.

3.2. Relationship between Microphone Spacing and NRR

The relationship between the spacing of paired-microphones and NNR in each frequency bin was examined by computer simulation. In Fig. 1, an objective sound source is located near the front of paired-microphones, and 14 interfering sources are located from 20 degrees to 80 degrees at both sides. Each interfering source generates a 1/3 octave-band noise whose central frequency changes from 125 Hz to 8000 Hz. The spacing of paired-microphone is different, from 0.01 m to 0.20 m, at intervals of 0.01 m.

Figure 2 gives NRRs, which suggest how much noise is reduced, in microphone spacing and frequency. As we expected, it is obvious that the most suitable spacing of paired-microphones is different at each frequency. According to Eq. (8), when two microphones are arranged with spacing of 0.20 m, another microphone should be placed at the position of 0.07 m in Fig. 1. In other words, the spacings of the paired-microphones are 0.20 m, 0.17 m, and 0.03 m.

3.3. Criterion on Paired-microphone Selection

The arrangement of a 3-ch non-equally-spaced microphone array is fixed in a condition in Fig. 1. NRR might change depending on both geometrical relationship and acoustic characteristics among sound sources, and does not take distortion of the objective signal into account. Accordingly, a criterion is introduced to select the most suitable paired-microphone at each frequency.

If objective signals are considered speech signals, the smoothness on estimated objective signals might be a clue for selecting the most suitable paired-microphone. An application is a noise reduction front-end for ASRs. Speech signals change smoothly in time because current ASRs analyze input signals every few decade seconds. On the other hand, this smoothness is not guaranteed in the spectral domain because of both resonance and anti-resonance caused by the vocal tract. Here, a cost function based on spectral smoothness is proposed by observing the amplitude spectrum. When signals received by three microphones are cut out by the temporal window with the k-th index by using signals $x_{k,i}(t)$ and $x_{k,j}(t)$ $(i, j = 1, 2, 3, \quad i \neq j)$, an estimated amplitude spectrum $\widehat{|S|}_{k,ij}(\omega)$ of the objective signal is obtained for each paired-microphone and each frequency. The criterion $D_{ij}(\omega)$ is simply defined as the difference between the current estimation $\widehat{|S|}_{k,ij}(\omega)$ for the ij-th paired-microphone and the most suitable estimation $\widehat{|S|}_{k-1}(\omega)$ for all paired-microphones in the previous frame.

$$D_{ij}(\omega) = \left| \widehat{|S|}_{k-1}(\omega) - \widehat{|S|}_{k,ij}(\omega) \right| \qquad (9)$$

Minimizing $D_{ij}(\omega)$ subject to every combination between i and j independently in each frequency ω, we can obtain temporally smooth spectra. It is easy to suppose that the criterion on spectral smoothness works well for vowels but its suitability is uncertain in rapidly changing sections such as the consonant part in a VCV sequence.

4. Performance Evaluation

Noise reduction performance is evaluated by computer simulation from the following points of view. The first comparison is done between non-equally-spaced and equally-spaced microphone arrays, and the second is done between the criteria of NRR and spectral smoothness.

4.1. Evaluation Measure

The proposed method is a kind of spectral subtraction, so spectral distortion is adopted as an evaluation measure. Spectral distortion is calculated segmentally based on the Euclidean distance between log amplitude spectra as,

$$\frac{1}{K}\sum_{k=1}^{K}\sqrt{\frac{1}{W}\sum_{\omega=1}^{W}\left(P_k^{(eval)}(\omega) - P_k^{(ref)}(\omega)\right)^2}\,[\text{dB}], \qquad (10)$$

where $P_k^{(eval)}(\omega)$ and $P_k^{(ref)}(\omega)$ correspond to a signal to be evaluated and a reference signal that must be a clean objective signal in the k-th frame, respectively, and W and K mean the total numbers of frequency bins and frames for evaluation, respectively.

4.2. Comparison between Non-equally-spaced and Equally-spaced Arrays

Noise reduction performance for a non-equally-spaced array is compared with a previously proposed equally-spaced array on the condition shown in Fig. 1. In this comparison, the most suitable paired-microphone is selected based on NRR shown in Fig. 2. An objective signal is a male utterance */My name is John/* coming from the front of the array. Interfering signals are a wide-band noise (125 Hz - 6000 Hz) generated from the surrounding sound sources shown in Fig. 1

Figure 3 shows waveforms both of a clean speech signal and a noise-added speech signal as well as spectral distortions for a noise-added signal (dashed line), noise-reduced signals by the previous equally-spaced array (dotted line), and the proposed non-equally-spaced array (solid line). From Fig. 3, we can easily confirm the advantage of the non-equally-spaced array.

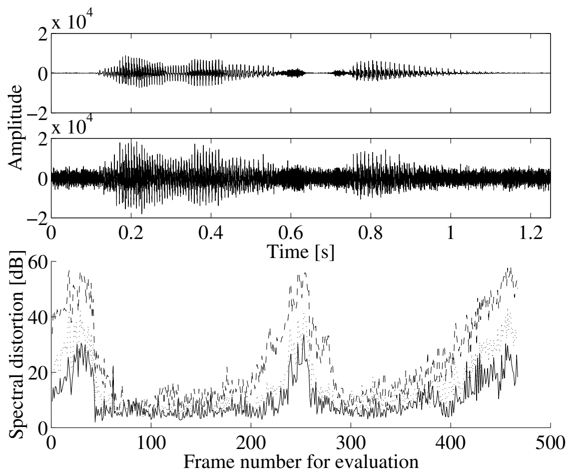

Figure 3: Waveforms of a clean speech uttered */My name is John/* and a noise-added speech, and spectral distortions for a noise-added speech (dashed line) and noise-reduced speech by using either equally-spaced array (dotted line) or non-equally-spaced array (solid line).

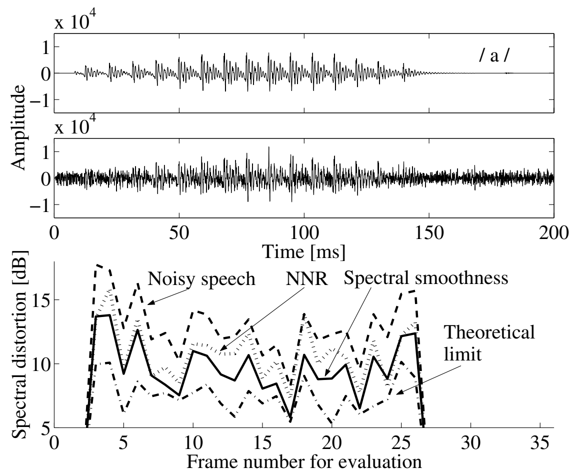

Figure 4: Waveforms of clean and noisy speech /a/, and spectral distortions for noise-added speech (dashed line), noise-reduced speech by NRR criterion (dotted line), by spectral smoothness criterion (solid line), and by theoretically optimized selection (dash-dot line).

4.3. Comparison of Critera for Paired-microphone Selection

This comparison focuses on criteria for selecting the most suitable paired-microphone at each frequency. The criteria used here are NRR defined by Eq. (7) and spectral smoothness defined by Eq. (9). Three microphones are arranged in non-equal spacing. Objective signals consist of speech that is assumed to come from near the front of an array, and an interfering signal is a wide-band noise coming from 30 degrees to the right.

First, Japanese vowel /a/ is used as a stationary objective signal. The upper area of Fig. 4 shows waveforms of clean and noise-added speech signals, and the lower area shows spectral distortions for a noise-added signal (dashed line), noise-reduced signals by NRR criterion (dotted line), those by spectral smoothness criterion (solid line), and theoretically optimized selection (dash-dot line) that is given by selecting the minimum spectral distortion between an estimated objective signal and a reference signal. Figure 4 shows further improvement by using the spectral smoothness criterion instead of NRR.

Next, the objective speech signals are 10 natural utterances, including rapid changes from the ATR spontaneous speech database. For every utterance, spectral distortion is calculated for noisy speech, noise-reduced speech by NRR criterion, or that by spectral smoothness criterion. The mean spectral distortions over 10 utterances are 13.2 dB, 8.5 dB, and 8.0 dB in order. Spectral smoothness is a suitable criterion not only for stationary signals but also for natural speech.

5. Conclusions

In this paper, a non-equally-spaced microphone array is proposed for noise reduction by paired-microphones with the aim of improving performance. The appropriate arrangement is decided based on a noise reduction rate for a 3-ch linear array. Assuming that objective signals are speech, spectral smoothness criterion is also introduced for selecting the most suitable paired-microphone. The feasibility of both the non-equally-spaced array and the criterion of spectral smoothness are confirmed by computer simulation. In the future, the proposed method should be evaluated in the real environment.

Acknowledgment

This research was supported in part by the Telecommunications Advancement Organization of Japan.

6. References

[1] J. C. Junqua and J. P. Haton, *Robustness in automatic speech recognition*, Kluwer Academic Publishers, Boston, 1996.

[2] S. F. Boll, "Suppression of acoustic noise in speech using spectral subtraction," IEEE Trans. ASSP, Vol. 27, No. 2, pp. 113-120, 1979.

[3] J. L. Flanagan, D. A. Berkley, G. W. Elko, J. E. West and M. M. Sondhi, "Autodirective microphone systems," ACUSTICA, Vol. 73, pp. 58-71, 1991.

[4] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*, Springer Verlag, New York, 2001.

[5] M. Mizumachi and M. Akagi, "Noise reduction by paired-microphones using spectral subtraction," Proc. ICASSP, Vol. II, pp. 1001-1004, 1998.

[6] M. Mizumachi and M. Akagi, "Noise reduction method that is equipped for a robust direction finder in adverse environments," Proc. IEEE Workshop on Robust Methods for Speech Recognition in Adverse Conditions, Tampere, Finland, pp. 179-182, 1999.

[7] C. H. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," IEEE Trans. ASSP, Vol. 24, No. 4, pp. 320-327, 1976.