Noise Reduction Using Paired-microphones for both Far-field and Near-field Sound Sources

Mitsunori Mizumachi and Satoshi Nakamura
ATR Spoken Language Translation Research Laboratories
E-mail: mitsunori.mizumachi@slt.atr.co.jp

Abstract
The near-field problem is a source of anxiety with beamforming techniques. We propose a strategy to solve the near-field problem by using a subtractive beamforming technique. The authors earlier proposed a method for noise reduction using paired-microphones. In this paper, the method is improved for near-field sound sources. The proposed method can maintain a high performance regardless of the distance between the sound source and an array, but the performance of a Delay-and-Sum beamformer declines even if the amplitude of the target signal is normalized. The concept of “paired-microphones” in the proposed method is the key for solving the near-field problem.

1. INTRODUCTION
Noise reduction research has a long history, and has recently been watched with keen interest to make automatic speech recognizers (ASRs) robust in the real world. It is well known that current ASRs do not work well under noisy environments [1]. The expectation for hands-free speech recognition has also accelerated demands for noise reduction techniques.

A wide variety of methods for noise reduction have been proposed to make ASRs robust, and can be broadly divided into two approaches. One comprises signal processing approaches such as a microphone array which is a front-end for ASRs, and the other comprises model adaptation approaches in the HMM framework such as the HMM decomposition technique [2]. Formerly, almost all methods aimed at reducing stationary noises or recovering from speech distortions caused by stationary noises. Spectral Subtraction (SS) [3] is a typical algorithm for reducing stationary noises, and is employed practically in many current ASRs. Unfortunately, the method is extremely weak for non-stationary noises, although there are many sources of non-stationary noises around ASRs. We therefore need to remove the effects of various real noises to make ASRs robust or achieve comfortable hands-free speech communications in the real world.

The authors earlier proposed an algorithm for noise reduction with a 3ch linear microphone array [4]. This method adopts a subtractive beamforming technique based on “paired-microphones”, and has a subtractive beamformer analytically designed without using an adaptive filter. An adaptive beamformer, for example, Griffiths-3m beamformer [5], works very well for a few specific noises under ideal environmental conditions. However, it is hard to deal with non-stationary noises, moving noises, and all signals received under reverberant environments, since an adaptive filter cannot converge under such conditions.

In place of using an adaptive filter, the spectra of interfering signals are analytically estimated on the basis of arrival time differences between paired-microphones frame by frame. Next, the estimated noise spectra are subtracted from noisy received signals in the frequency domain just like the SS. The proposed method can reduce non-stationary noises under not only a computer simulated environment, but also a real environment, because the noise spectra are correctly estimated in each short term frame [4]. The feasibility of the proposed method is confirmed for far-field sound sources as a front-end for ASRs under both noisy and reverberant environments [6].

In this paper, the authors intend to prove the availability of the method for near-field sound sources. Almost all beamformers assume that signals arrive at a microphone array as plane waves, that is, all sound sources are far from the array. It is difficult for hands-free speech communications to satisfy this assumption, because a PDA or a wearable PC must capture speech signals. Under near-field conditions, spherical waves make the noise reduction performance worse because the waves cause differences not only in the phase spectrum but also in the power spectrum of each signal. A Delay-and-Sum beamformer must normalize the power among all received signals, so it has to know the distance between the sound source and each microphone. The authors try to solve the near-field problem by using a subtractive beamformer that is designed by only the phase information between paired-microphones.

Results of performance tests show that the proposed method can deal with near-field sound sources with no trouble. The performance of the proposed method does not depend on the distance between the sound source and an array, because the proposed method requires only the arrival time differences between paired-microphones, that is, a unique cue for designing a subtractive beamformer. It is very important for the proposed method to use paired-microphones.

In section 2, we describe an outline of our formerly proposed algorithm for noise reduction [4][7]. In section 3, what we have to solve is clearly pointed out in the near-field problem. In section 4, we verify the feasibility of the proposed method under near-field conditions, and the proposed method is improved by taking advantage of characteristics for near-field conditions. Finally, the conclusion is given in section 5.
2. ALGORITHM DESCRIPTION

The authors earlier proposed an algorithm for noise reduction using a 3-ch linear microphone array. The microphone array consists of three linearly arranged microphones at intervals of 10 cm. This method works by combining spatial filtering and frequency filtering.

The spatial filter is analytically constructed based on signal directions. To put it into practice, we proposed a subtractive beamformer [4] and a robust direction finder under noisy environments [5]. Here, we give outlines of them and an overview of noise reduction.

2.1. Subtractive Beamformer [8]

The proposed subtractive beamformer is designed using two signals received by paired-microphones, for the purpose of eliminating directional noise signals. Let us assume that a target signal \( s(t) \) and an obstructive noise signal \( n(t) \) exist. We can obtain the signal received by each microphone as follows.

\[
\text{left mic.:} \quad l(t) = s(t - \zeta) + n(t - \delta),
\]

\[
\text{center mic.:} \quad c(t) = s(t) + n(t),
\]

\[
\text{right mic.:} \quad r(t) = s(t + \zeta) + n(t + \delta),
\]

where \( \zeta \) means the arrival time difference between the neighboring microphones for target signal \( s(t) \), and \( \delta \) means the difference for noise signal \( n(t) \).

If we know that \( \delta \) equals the direction of the noise signal in advance, the beamformer \( g_i(t) \) can completely eliminate noise signal \( n(t) \).

\[
g_i(t) = \frac{1}{2} \left\{ \left[ l(t + \delta + \tau) - l(t + \delta - \tau) \right] - \left[ r(t - \delta + \tau) - r(t - \delta - \tau) \right] \right\},
\]

where \( \tau \) is an arbitrary constant except zero. It is sufficient to use only a single temporal shift \( +\tau \) or \(-\tau \) to eliminate the noise signal. Taking account of both \(+\tau \) and \(-\tau \), we can obtain the noise eliminated signal at the position of the center microphone.

Then the short-term Fourier transformation \( G_i(\omega) \) of beamformer \( g_i(t) \) can be calculated as

\[
G_i(\omega) = S(\omega) \sin \omega(\zeta - \delta) \sin \omega \tau.
\]

It is important to note that \( G_i(\omega) \) has no term concerned with noise signal \( n(t) \). Setting \( \tau \) as \( \zeta - \delta \), the spectrum \( S(\omega) \) of target signal \( s(t) \) can be estimated.

\[
S(\omega) = G(\omega) / \sin^2 \omega(\zeta - \delta),
\]

\[
\omega(\zeta - \delta) \neq n \tau \quad (n: \text{integer}.
\]

Here, the accuracy in estimating noise spectra may decrease in specific regions defined by both signal directions and the signal frequency.

2.2. Direction Finder [7]

The direction finder used to construct the subtractive beamformer must work well under noisy environments. The authors succeeded in making the direction finder robust by integrating the subtractive beamformer with a traditional cross-correlation method.

First, the direction of the most dominant signal is estimated based on the whitened cross-correlation between \( l(t) \) and \( r(t) \). It turns out that the signal coming from this direction is either the target signal or the noisiest signal which should be reduced.

Next, we obtain two signals that suppress the above signal completely using the subtractive beamformer. These signals are equal to beamformer \( g_i(t) \), calculated with Eq. (4) using \( l(t) \) and \( c(t) \), and beamformer \( g_r(t) \) as well. The second direction is also estimated based on the whitened cross-correlation between \( g_i(t) \) and \( g_r(t) \).

Finally, we must decide which directions correspond to the target signal and the noise signal in each frame. In the first frame, assuming that the target signal comes nearer to the front than the noisiest signal, we can distinguish the two signal directions. From the second frame onward, each direction estimated in the preceding frame becomes an index. The sound source moves smoothly in ordinary circumstances. Accordingly, we can obtain the directions of both the target signal and the noise signal in each frame.

2.3. Outline of Noise Reduction

The noise reduction procedure is as follows. In each short term frame, after the two signal directions are found, the noise spectrum is estimated by choosing the most suitable beamformer automatically in each frequency bin. The noise reduction is rounded out by subtracting the noise spectrum from the spectrum of the received signal, i.e., non-linear SS.

3. NEAR-FIELD PROBLEM

The near-field problem gives a blow to the conventional beamforming theory. The conventional Delay-and-Sum beamformer assumes that the target signal arrives at an array as a plane wave. The target sound source is far from the array in other words. The differences among the powers of received signals can generally be ignored under far-field conditions. Under near-field conditions, however, spherical waves come to the array. The above assumption cannot be satisfied as a result, because the near-field sound source causes differences in not only the phase spectrum but also in the power spectrum for each received signal.
3.1 Formulation of near-field sound sources

Let us assume that the distance between a sound source that presents a signal $s(t)$ and the central microphone of a 3ch linear microphone array is $d$, and signal $s(t)$ comes from the direction $\theta$ as shown in Fig. 2. Here, the distances between the sound source and each microphone can be defined as follows.

\[ l_i = \sqrt{d^2 + \frac{d^2}{2} - 2dl_i \cos(\theta + \frac{\pi}{2})} \]

\[ l_i = \sqrt{d^2 + \frac{d^2}{2} - 2dl_i \cos(\theta - \frac{\pi}{2})} \]  \hspace{1cm} (8)

In the free-field, i.e., a complete anechoic room, the amplitude of a signal decreases in inverse proportion to the square of the distance by the inverse-square law. In the diffuse-field, i.e., a complete reverberation room, the amplitude of a signal does not decrease at all. When we think about the acoustics characteristics of a typical room in daily life, they should be intermediate between those of the free-field condition and the far-field condition. In this paper, therefore, we define the signal received by each microphone as follows.

\[ z_i(t) = \frac{1}{l_i} s(t - \frac{l_i}{c}), \quad i = 1, 2, 3 \]  \hspace{1cm} (9)

4. EXPERIMENTS AND RESULTS

4.1 Verification of the proposed method

In this section, the feasibility of the proposed method is verified under near-field conditions. We also discuss the advantage of the proposed method compared with a Delay-and-Sum beamformer under near-field conditions.

To examine the beam-pattern of each beamformer exactly, we assume that the signal direction is known for all beamformers. The beam-pattern is shown as an average pattern from 125Hz to 6kHz, when the target signal comes from the front of an array and an interference signal comes from 60 degrees to the right. The interval between neighboring microphones is 10cm, the sampling frequency is 48 kHz, and the distance between the sound source and the central microphone of the 3ch linear array is 1.0m, 0.30m, or 0.15m. Beam-patterns for distances of 1.0m, 0.30m, and 0.15m, are shown in Fig. 3, Fig. 4, and Fig. 5 respectively. We use two types of Delay-and-Sum beamformers. One is the conventional Delay-and-Sum beamformer which knows only the signal direction, and the other is an amplitude-normalized Delay-and-Sum beamformer that knows both the signal direction and the distances between the sound source and each microphone.

In Fig. 3, the beam-pattern of the conventional Delay-and-Sum beamformer represented by "x“, equals that of the amplitude-normalized Delay-and-Sum beamformer represented by "Δ". Therefore, the far-field assumption is applicable when the distance is 1.0m. However, as the sound source approaches the array, the performance levels of both types of Delay-and-Sum beamformers become worse, even if the amplitude of the target signal can be normalized exactly, because the far-field assumption is broken. On the other hand, the performance of the proposed method does not decline at all even under near-field conditions. It is sufficient for the proposed method to know the arrival time differences between paired-microphones. They are only cues for designing the subtractive beamformer defined in Eq. (4). The concept of "paired-microphones" is the key to maintaining a high performance regardless of the distance between the sound source and an array. Furthermore, the proposed method does not care whether a signal comes as a plane wave or not.

4.2 Improvement of the proposed method

To improve the performance of the proposed method, we reconsider the near-field conditions. Under far-field conditions, the arrival time differences depend only on the distance between the paired-microphones, because the signal direction is considered to be the same for all paired-microphones. Under near-field conditions, the signal direction may differ between the paired-microphones. The former proposed method used only two paired-microphones, consisting of microphones at both ends or consisting of a center microphone and a left microphone, but we can set three available paired-microphones under near-field conditions. The improved method selects the most suitable paired-microphones at each frequency.
based on the computational accuracy, that is, the value of the denominator in Eq. (6).

The feasibility of the improved method can be examined objectively by the noise reduction rate (NRR).

\[
NRR = 10 \log_{10} \frac{1}{W} \sum_{\omega=1}^{W} \frac{|N_{\text{input}}(\omega)|^2}{|N_{\text{output}}(\omega)|^2} \ [\text{dB}] \quad (10)
\]

where \(W\) is the upper limit to evaluate in frequency, and \(N_{\text{input}}(\omega)\) and \(N_{\text{output}}(\omega)\) are the spectra of the noise signal before and after noise reduction. Figure 6 shows the ability of noise reduction in each direction of a noise signal (random noise from 125 Hz to 6 kHz) for both the improved proposed method and the former proposed method. The source of the noise signal is 50 cm away from the center microphone of a 4-channel linear array. This figure indicates that the improvement by using three paired-microphones is good under near-field conditions.

**5. CONCLUSION**

In this paper, we have discussed the near-field problem with beamforming techniques. The proposed method can maintain a high performance under even near-field conditions, but the performance of a Delay-and-Sum beamformer significantly drops even if the amplitude of the target signal is normalized. The concept of “paired-microphone” is the key for solving the near-field problem in the proposed method.

In the future, the authors plan to build a robust direction finder for use in adverse environments, and to examine the performance of the proposed method in the real world.

**6. References**


