Noise reduction method for multiple sound source signals and its application to two-channel microphone system

Takaaki Ishibashi*a,+, Kana Higuchia, Chiharu Okumaa, Kiyoteru Hayamaa

aNational Institute of Technology, Kumamoto College
2659-2 Suya, Koshi, Kumamoto 861-1102, Japan

*Corresponding Author: ishibashi@kumamoto-nct.ac.jp

Abstract

This paper proposes a noise reduction method for multiple sound source signals. The proposed method can change the direction of unidirectional pattern and can select multiple directivity pattern. The method can reduce noise even when the number of sound sources is larger than that of observed mixture signals. In addition, a noise reduction system is developed using two-channel microphone arrays. The system can be processed in real-time. By several simulations, it is found that the proposed method can extract the target speech signals from multi speaker’s utterance sounds.

Keywords: noise reduction, variable arbitrary directivity pattern, two-channel microphone system.

1. Introduction

Many noise reduction methods such as ICA (Independent Component Analysis)1,2,3, SS (Spectral Subtraction)4, SAFIA (sound source Segregation based on estimating incident Angle of each Frequency component of Input signals Acquired by multiple microphones)5 and NMF (Nonnegative Matrix Factorization)6 have been proposed. The methods can estimate the original source signals. However, these algorithms are complicated and take much time to calculate in applications using a micro controller.

In this paper, we propose a simple two-channel microphone system with variable arbitrary directional pattern. A typical microphone forms unidirectional directivity pattern. Therefore, when a multiple speaker speaks at the same time, it is necessary to place as many microphones as the number of speakers. Our method works well even when the number of sound sources is larger than that of microphones. The target speech of one person or a plural person is emphasized by our method. And another speech and noise are reduced. Since the proposed method is a very simple algorithm and the amount of calculation is small, it can be expected to be applied to industries.

2. Control of directional pattern

2.1 Estimation for direction of source

When some sound sources $s_n(t) \,(n = 1, 2, \ldots, N)$ are observed by some microphones, the observed mixture signals $x_m(t) \,(m = 1, 2, \ldots, M)$ are expressed as follows.

$$x_m(t) = \sum_{n=1}^{N} a_{mn} s_n(t) \quad (1)$$

where $a_{mn}$ denotes a mixing parameter, $t$ denotes index of time series, $N$ and $M$ are the number of sound sources and mixture signals, respectively.

Using two mixture signals observed by two microphones in the case of two speaker’s utterances at the same time, a joint distribution is plotted as Fig. 1(a). The horizontal and vertical axes denotes the amplitude of $x_1(t)$ and $x_2(t)$, respectively. In the following, we consider the audio signals as a stochastic variety and omit the time series $t$.

From the Fig. 1(a), it is found that the joint distribution has two linear components. For clarity the linear components, directions $\phi$ of distribution are calculated as follows.

$$\phi = \tan^{-1} \frac{x_2}{x_1} \quad (2)$$

Using all $\phi$, its histogram $\text{hist}(\phi)$ as shown in Fig. 1 (b) means that the components of the joint distribution are concentrated two directions. The reason is that since the human speech has a silent interval, it becomes only one source in many data points.
In the case of three active sources, a joint distribution and a histogram observed two microphones are shown in Fig. 2(a) and (b), respectively. In these figures, the dense crossing lines are still discernible and three peaks are recognizable. These values of the inclinations of lines depend on the transfer functions from the sources to the microphones.

\[
\Phi = \arg\max_{\phi} \text{hist}(\phi)
\]  
(4)

In the case which the target source signal is multiple speakers speech, we estimate another peak \(\Phi_k\) again excluded the mode value \(\Phi\) and neighbor value.

In order to reduce sounds in a direction away from the target source, we use some functions \(f_i(\varepsilon)\) that gain decreases

\[
\varepsilon = \Phi - \phi
\]  
(3)

where \(\Phi\) denotes the direction of the target source. From the above discussion, \(\Phi\) is estimated as the mode value (the most frequent value) of \(\phi\) as follows.

To Extract the target speech using the joint distribution means that to reduce the straight line component generated by the noise, and to extract the straight line components created by the target signal. Based on the fact, the authors have proposed a noise reduction method based on the ratio of mixture signals observe two microphones\(^7,8\). We define the error \(\varepsilon\)

\[
\varepsilon = \Phi - \phi
\]  
(3)

where \(\Phi\) denotes the direction of the target source. From the above discussion, \(\Phi\) is estimated as the mode value (the most frequent value) of \(\phi\) as follows.

In the case which the target source signal is multiple speakers speech, we estimate another peak \(\Phi_k\) again excluded the mode value \(\Phi\) and neighbor value.

In order to reduce sounds in a direction away from the target source, we use some functions \(f_i(\varepsilon)\) that gain decreases
as the $\varepsilon$ increases as

$$f_1(\varepsilon) = \frac{1}{1 + \alpha \varepsilon^2}$$ \hspace{1cm} (5)

$$f_2(\varepsilon) = -0.5 \tanh(\beta \varepsilon^2 - \gamma) + 0.5$$ \hspace{1cm} (6)

$$f_3(\varepsilon) = \begin{cases} 1 & \text{if } \varepsilon^2 \leq \delta \\ 0 & \text{if } \varepsilon^2 > \delta \end{cases}$$ \hspace{1cm} (7)

where $\alpha$, $\beta$, $\gamma$ and $\delta$ denote learning parameters for determining the amount of noise reduction. Fig. 3 (a), (b) and (c) show orthogonal coordinates and polar coordinates of $f_1(\varepsilon)$, $f_2(\varepsilon)$ and $f_3(\varepsilon)$, respectively. From these figures, the noise reducing functions $f_i(\varepsilon)$ mean spatial filters on the ratio of the transfer functions.

The target signal $y$ is estimated based on the $\phi$ using observed signals $x_m$ mixed multiple source signal as follows.

$$y = f_i(\varepsilon)x_m$$ \hspace{1cm} (8)

3. Variable arbitrary directivity pattern

3.1 Directivity in multiple directions

The noise reduction functions proposed by the authors has been unidirectional. It means that the functions $f_i(\varepsilon)$ emphasized only one sound. In some applications, it is often that the target signal is a speech of multiple speakers. Therefore, we propose a method that can record multiple speaker’s speech reduced other noise.

As a function with multiple directivity pattern, we chose the Rose curve $g_l$ as follows.$^9$

$$g_l = \begin{cases} \sin(l(\theta - \varphi)) & \text{if } l : \text{odd} \\ \sqrt{\sin(l(\theta - \varphi))} & \text{if } l : \text{even} \end{cases}$$ \hspace{1cm} (9)

where $\theta$ denotes the angle from 0 to 360 degrees and $\varphi$ denotes the phase. The Rose curves is shown in Fig. 4. From the figures, the directivity can be changed shape and direction by using $g_l$.

3.2 Noise reduction system

Using the noise reduction functions $f_i(\varepsilon)$ and the Rose curve $g_l$, a variable arbitrary directivity pattern is proposed as follows and some directivity patterns are shown in Fig. 5.

$$y_m(t) = g_l f_i(\varepsilon)x_m$$ \hspace{1cm} (10)

Even when the number of sound sources is larger than the number of observed mixture signals, our proposed method can reduce noise signals. In addition, the algorithm of our noise reduction method is very simple. Then the method works in real-time. From the same way, the method can also

reduce only one source signal. It means that the method has a function to reduce only the target source signal.

4. Simulation

In order to verify our proposals, several simulations were carried out. The target source signal $s_1(t)$ was human speech signal in the database.$^{10}$ The noise source $s_2(t)$ was another human speech. Fig. 6 shows the source signals. These signals were sampled at the rate of 8 [kHz] with 16 [bit] resolution.

In order to clarify the speech segment, $s_1(t)$ was active from 1 to 6 [s] and $s_2(t)$ is active from 4 to 7 [s]. Using these sources, the mixture signals are generated by Eq (1) with the mixing parameters $a_{mn} = 0.9 \pm \eta$ $(m = n)$ and $a_{mn} = 0.6 \pm \eta$ $(m \neq n)$, $\eta$ denotes a random value from 0 to
The mixture signals are shown in Fig. 7. It is confirmed that the mixture signals have amplitude in 1 to 7 [s].

Fig. 7: Waveforms of observed signals

The estimate signal using by Eq (5) and $\alpha = 0.5$ is shown in Fig. 8. From simulation result, it is found that the target source signal $s_1(t)$ is extracted and the noise signal $x_2(t)$ is reduced by the proposed method.

We confirmed that the method can obtained similar results with other combination of speech data (30 mixture patterns: permutation of 3 males and 3 females $P_3$) and some stationary noise (30 mixture patterns: 6 human speech and 5 car noise). It is confirmed that the proposed method works well even when the number of sound sources increases. Similar results were obtained in other functions $y(t)$.

Fig. 8: Waveform of estimated signal

5. Conclusion

The two-channel microphone system with the variable arbitrary directivity pattern is proposed. The proposed method is reduced the noise by deleting the distribution except a linear component depending on the target speech. The proposed method can reduce the noise even when the number of source signals is larger than that of observed mixture signals. The method is very simple and its application with microcontroller can be expected.

References