Directionally Constrained Minimization of Power Algorithm for Speech Signals

a)Takahiro Murakami, b) Kiyoshi Kurihara, and c) Yoshihisa Ishida

Department of Electronics and Communication, Meiji University, Japan

a)tmura@isc.meiji.ac.jp
b)ce44031@isc.meiji.ac.jp
c)ishida@isc.meiji.ac.jp

Abstract

A novel algorithm of adaptive beamforming with a microphone array is presented. Beamforming is one of the simplest methods for discriminating between different signals based on the physical location of sources. The directionally constrained minimization of power (DCMP) algorithm is well-known as a reliable approach to extracting some signals incoming from specific directions. In the DCMP algorithm, the output power of the sensor array is minimized under the constraint of a constant response to the direction of arrival (DOA) of the target signal. Although the conventional DCMP is only applied to narrowband signals, our approach extends this algorithm so that the DCMP criterion is applicable to the broadband signals such as speech. The proposed method can be achieved by a relatively small hardware amount of the system.

1. Introduction

To apply various speech applications, e.g., automatic speech recognizers, hands-free mobile telephony, and hearing aids for impaired people, extraction of desired speech from observations at microphones, which contain the target speech, undesirable speech by other talkers, background music and so on, is required. When multichannel sensor is available, adaptive beamforming is one of the simplest methods of separating the desirable speech by other talkers, background music and so on, at microphones, which contain the target speech, undesirable sources, respectively. In the adaptive beamformer, a multiple-input canceller, and a blocking matrix is generally carried out by delay-and-sum and delay-and-subtract operations, respectively, the GJBF requires a relatively large hardware amount of the system. The Griffiths-Jim beamformer (GJBF) is the most widely known method as the adaptive beamformer [4]. In the GJBF, the target signals are retrieved by minimizing the power of array output under a linear constraint. The aim of the linear constraint is to avoid zero output. In general, the GJBF is composed of a fixed beamformer, a multiple-input canceller, and a blocking matrix. In addition, since the fixed beamformer and the blocking matrix is generally carried out by delay-and-sum and delay-and-subtract operations, respectively, the GJBF requires delay systems as pre-processing. The delay systems are employed so that the components of the specific signal incoming from the direction of the target source, are in phase. Therefore, the GJBF is constructed by a relatively large hardware amount of the system.

Alternatively, the directionally constrained minimization of power (DCMP) algorithm would be available if the signals are narrowband [5]. The DCMP algorithm is exploited in the array antennas. In the DCMP algorithm, under a linear constraint based on a directional response, the power of output of the sensor array is minimized. Since the DCMP algorithm needs no delay systems as pre-processing, it can be realized by a relatively small hardware amount of the system. The conventional DCMP algorithm is, however, not applicable to speech signals since speech has high-bandwidth frequency components.

In this paper, we present a novel algorithm of adaptive beamforming for speech signals. Our method is essentially based on the DCMP criterion. By extending the DCMP criterion so that it can be applied to broadband signals, a beamformer with simple structure, in which only an FIR filter at each sensor is adaptively modified to form its directivity pattern, is derived.

2. The model

Let discrete-time observations at a K-channel sensor array be \( x_k(t) \) \((k = 1, 2, \ldots, K)\). When \( x_k(t) \) is composed of \( M(\leq K) \) source signals, \( x_k(t) \) can be modeled as

\[
x_k(t) = \sum_{m=1}^{M} s_m(t - \tau_m(k)),\quad (k = 1, 2, \ldots, K),
\]

where \( s_m(t) \) \((m = 1, 2, \ldots, M)\) is the \( m \)-th source signal incoming from the direction \( \theta_m \), and \( \tau_m(k) \) denotes the time delay in \( s_m(t) \). In this paper, for the sake of simplicity of explanation, we assume that the sensor array is an uniform linear array in which distance between adjacent sensors is \( d \). Therefore, \( \tau_m(k) \) is given by

\[
\tau_m(k) = \frac{f_s (k - 1)d \sin \theta_m}{c},\quad (k = 1, 2, \ldots, K; m = 1, 2, \ldots, M),
\]

where \( f_s \) and \( c \) indicate the sampling frequency and velocity of signals, respectively. Note that \( \tau_m(k) \) can be negative when the signal is incoming from the left side of the sensor array, i.e., \( \theta_m < 0 \). Figure 1 shows an example of a \( K \)-channel sensor array.

3. DCMP algorithm for narrowband signals [5]

Let the signals be narrow-bandwidth with the normalized center frequency \( f_c \) \((0 \leq f_c < 1)\). In the adaptive beamformer, the output of the sensor array is generated by summing up the...
adjusted forming, restoration of the source signals is carried out by adaptively constraint is employed in order to avoid zero output of the undesirable signals. The DCMP algorithm is defined by

\[
\begin{align*}
\mathbf{w} &= [w_1, w_2, \cdots, w_K]^T, \\
\mathbf{x}(t) &= [x_1(t), x_2(t), \cdots, x_K(t)]^T.
\end{align*}
\]

The role of \( w_k \) is to modify both the amplitude and phase of \( x_k(t) \).

The adaptive beamforming is one of algorithms for retrieving specific source signals from \( x_k(t) \). In the adaptive beamforming, restoration of the source signals is carried out by adjusting \( \mathbf{w} \) so as to direct null of the sensor array toward the DOAs of the undesirable signals. The DCMP algorithm is performed by minimizing the power of \( y(t) \) under the constraint of a constant response to the DOAs of signals. This directionally constraint is employed in order to avoid zero output of the sensor array. The DCMP algorithm is defined by

\[
\min_{\mathbf{w}} \left( P_{\text{out}} = \frac{1}{2} \mathbf{w}^H \mathbf{R}_{xx} \mathbf{w} \right)
\]

subject to

\[
\mathbf{C}^T \mathbf{w}^* = \mathbf{h},
\]

where

\[
\mathbf{C} = [\mathbf{c}_1, \mathbf{c}_2, \cdots, \mathbf{c}_N], \\
\mathbf{h} = [h_1, h_2, \cdots, h_N]^T.
\]

\( \mathbf{R}_{xx} \) is an autocorrelation matrix of \( \mathbf{x}(t) \) obtained as

\[
\mathbf{R}_{xx} = E[\mathbf{x}(t)\mathbf{x}^H(t)],
\]

where \( E[\cdot] \) is an expectation operation. \( \mathbf{c}_n \) and \( h_n \) (\( n = 1, 2, \cdots, N \)) are referred to as the constraint vector and constraint response, respectively. \( \mathbf{c}_n \) is described in the form

\[
\mathbf{c}_n = [c_{1n}, c_{2n}, \cdots, c_{Kn}]^T, \quad (n = 1, 2, \cdots, N),
\]

where

\[
c_{kn} = e^{-j2\pi f_n t_n},
\]

\[
\tau_{kn} = \frac{f_s(k - 1)d \sin \phi_n}{c},
\]

\( k = 1, 2, \cdots, K; n = 1, 2, \cdots, N \),

(10)

and \( \phi_n \) is a constraint direction in radian. In general, \( h_n \) is set to either unity when \( \phi_n \) is associated with the target signal, or zero when \( \phi_n \) is corresponding to the undesirable signal. The optimum solution of (5) under the constraint (6) is obtained as

\[
\mathbf{w}_{opt} = \mathbf{R}_{xx}^{-1} \mathbf{C}(\mathbf{C}^H \mathbf{R}_{xx}^{-1} \mathbf{C})^{-1} \mathbf{h}^*.
\]

In (11), an estimate of \( \mathbf{R}_{xx}^{-1} \) is required. Practically, however, it is not available since \( x_k(t) \) has finite data length. Instead of (11), actually, the adaptive beamformer is employed, which modifies \( \mathbf{w} \) sample-by-sample using some leaning rule. One of the leaning algorithms is based on the least mean square (LMS) algorithm. The LMS algorithm for the DCMP criterion is given by

\[
\hat{\mathbf{w}}(t + 1) = \mathbf{P}[\hat{\mathbf{w}}(t) - \mu \mathbf{x}(t) y^*(t)] + \mathbf{g},
\]

\[
\hat{\mathbf{w}}(0) = \mathbf{g},
\]

(12)

where

\[
\mathbf{P} = \mathbf{I} - \mathbf{C}(\mathbf{C}^H \mathbf{R}_{xx}^{-1} \mathbf{C})^{-1} \mathbf{C}^H,
\]

\[
\mathbf{g} = \mathbf{C}(\mathbf{C}^H \mathbf{R}_{xx}^{-1} \mathbf{C})^{-1} \mathbf{h}^*.
\]

(13)

and \( \hat{\mathbf{w}}(t) \) is the estimate of \( \mathbf{w} \) at the \( t \)-th sample and \( \mu \) denotes the step-size parameter. Alternatively, the sample matrix inversion (SMI) algorithm can also be available. The SMI algorithm estimates \( \mathbf{R}_{xx}^{-1} \) sample-by-sample, and then, \( \hat{\mathbf{w}} \) is estimated by (11). The SMI algorithm for the DCMP criterion is described as

\[
\hat{\mathbf{R}}_{xx}^{-1}(t) = \frac{1}{\beta} \hat{\mathbf{R}}_{xx}^{-1}(t - 1) - \frac{(1 - \beta)\hat{\mathbf{R}}_{xx}(t - 1)\mathbf{x}(t)\mathbf{x}^H(t)\hat{\mathbf{R}}_{xx}^{-1}(t - 1)}{\beta^2 + \beta(1 - \beta)\mathbf{x}^H(t)\hat{\mathbf{R}}_{xx}(t - 1)\mathbf{x}(t)},
\]

(14)

where \( \hat{\mathbf{R}}_{xx}^{-1}(t) \) is an estimate of \( \mathbf{R}_{xx}^{-1} \) at the \( t \)-th sample and \( \beta \) is the step-size parameter as \( 0 < \beta < 1 \).

4. Proposed method

In this section, a novel method of adaptive beamforming is presented. Our algorithm extends the DCMP criterion to broadband signals. Fig.2 shows a block diagram of our algorithm. As shown in this figure, the input signals are filtered by the adaptive filter (ADF), and then, the filtered signals are summed up for the array output. In this paper, the finite impulse response (FIR) filter is considered.

In our method, the filter coefficients are optimized so that the power of the array output, say \( y(t) \), is minimized. To avoid the zero output, the adaptive learning algorithm must be constrained under some criterion. In this paper, a novel constraint for broadband signals is derived by extending the conventional DCMP algorithm.

Consider a broadband signal such as speech. The broadband signal can be regarded as consisting of a number of narrowband signals as in the form

\[
x(t) = \sum_{f} x(f; t),
\]

(15)
where $x(f; t)$ is the narrowband signal with the normalized center frequency $f$. Therefore, the constraint based on the DCMP criterion should be taken into account at each frequency. Then (6) is rewritten as

$$
\mathbf{C}^T(f)\mathbf{w}(f) = \mathbf{h}(f), \quad (n = 1, 2, \cdots, N). \tag{16}
$$

where $\mathbf{C}(f)$, $\mathbf{w}(f)$, and $\mathbf{h}(f)$ are the constraint matrix, the weight vector, and the constraint response vector at the frequency $f$, respectively. From (16), it can be said that this is the frequency-domain constraint. Therefore, by transforming (16) into the time domain, the DCMP criterion is extended to broadband signals.

Consider the $n$-th constraint. In the same way to (16), the $n$-th constraint at the frequency $f$ is given by

$$
e^{nT}(f)w^*(f) = \sum_{k=1}^{K} c_{kn}(f)w_k^*(f) = h_n(f). \tag{17}
$$

Since each term of (17) can be regarded as the multiplication of two spectra, i.e., $c_{kn}(f)$ and $w_k^*(f)$, the time-domain constraint is expressed by applying the inverse discrete-time Fourier transform (IDTFT). The IDTFT of (17) is described as follows:

$$
\sum_{k=1}^{K} \tilde{c}_{kn}(t)\tilde{w}_k(t) = \sum_{k=1}^{K} \sum_{p=-\infty}^{\infty} \tilde{c}_{kn}(t-p)\tilde{w}_k^*(p)
= \sum_{k=1}^{K} \sum_{p=0}^{P} \tilde{c}_{kn}(t-p)\tilde{w}_k^*(p)
= \sum_{k=1}^{K} \tilde{c}_{kn}(t)\tilde{w}_k^*
= \tilde{h}_n(t), \tag{18}
$$

where $\tilde{c}_{kn}(t)$, $\tilde{w}_k(t)$, and $\tilde{h}_n(t)$ are the IDTFT of $c_{kn}(f)$, $w_k(f)$, and $h_n(f)$, respectively, and $*$ denotes the convolution operation. $\tilde{c}_{kn}(t)$ and $\tilde{w}_k(t)$ are respectively obtained as

$$
\tilde{c}_{kn}(t) = [\tilde{c}_{kn}(0), \tilde{c}_{kn}(t-1), \cdots, \tilde{c}_{kn}(t-P)]^T,
\tilde{w}_k = [\tilde{w}_k(0), \tilde{w}_k(1), \cdots, \tilde{w}_k(P)]^T. \tag{19}
$$

Note that, in (18), the convolution operation is truncated to $0 \leq t \leq P$. Therefore, the constraint for the broadband signals is obtained as in the form

$$
\mathbf{C}^T \mathbf{w}^* = \mathbf{h}, \tag{20}
$$

where

$$
\mathbf{C} = \begin{bmatrix}
\mathbf{C}_{11} & \cdots & \mathbf{C}_{1N} \\
\vdots & \ddots & \vdots \\
\mathbf{C}_{K1} & \cdots & \mathbf{C}_{KN}
\end{bmatrix},
\mathbf{C}_{kn} = [c_{kn}(-Q), \cdots, c_{kn}(Q)],
\mathbf{w} = [w_1^T, \cdots, w_K^T]^T,
\mathbf{h} = [h_1^T, \cdots, h_N^T]^T,
\mathbf{h}_n = [\tilde{h}_n(-Q), \cdots, \tilde{h}_n(Q)]^T. \tag{21}
$$

Since $c_{kn}(f)$ is given by (10), $\tilde{c}_{kn}(t)$ is equivalent to the impulse response of the time-delay system in which the filter input is delayed $\tau_{kn}$ samples. Consequently, we have

$$
\tilde{c}_{kn}(t) = \frac{1}{\pi(t - \tau_0 - \tau_{kn})} \sin (\pi(t - \tau_0 - \tau_{kn})), \tag{22}
$$

where $\tau_0 = P/2$ is a constant delay so as to shift the current input to the center of the filter. The aim of inserting $\tau_0$ is to alleviate the effect of the causality. Note that, in (22), $\tilde{c}_{kn}(t) = 1$ when $t - \tau_0 - \tau_{kn} = 0$. In addition, $h_n(f)$ is generally a constant value $h_n$ for any $f$. Therefore, $h_n$ is given by

$$
\tilde{h}_n(t) = \begin{cases}
h_n, & (t = 0) \\
0, & (t \neq 0).
\end{cases} \tag{23}
$$

On the other hand, $\tilde{w}_k$ represents the FIR filter at the $k$-th channel. Therefore, the array output $y(t)$ is obtained as

$$
y(t) = \sum_{k=1}^{K} \tilde{w}_k^T \tilde{a}_k(t) = \tilde{W}^T \tilde{X}(t), \tag{24}
$$

where

$$\tilde{a}_k(t) = [x_k(t), x_k(t-1), \cdots, x_k(t-P)]^T,
\tilde{X} = [\tilde{a}_1^T(t), \tilde{a}_2^T(t), \cdots, \tilde{a}_K^T(t)]^T. \tag{25}
$$

Hence, the power of $y(t)$ has the form

$$E[|y(t)|^2] = \tilde{W}^T \tilde{R}_{XX} \tilde{W}, \tag{26}
$$

where

$$\tilde{R}_{XX} = E[\tilde{X}(t)\tilde{X}^H(t)]. \tag{27}
$$

Finally, we get the following cost function:

$$\min_{\tilde{W}} \left( P_{out} = \frac{1}{2} \tilde{W}^T \tilde{R}_{XX} \tilde{W} \right) \tag{28}
$$

subject to (20).

The optimum solution of (28) under the constraint (20) is obtained as

$$\tilde{W}_{opt} = \tilde{R}_{XX}^{-1} \tilde{C} (\tilde{R}_{XX}^{-1} \tilde{C})^{-1} \tilde{H}^*. \tag{29}
$$

Since the structure of (29) is similar to the case of the conventional DCMP algorithm (11), the adaptive learning algorithm given by (12)-(14) are applicable.
5. Simulation

In this simulation, we assumed that there are two sound sources and two microphones as \( K = M = 2 \). In addition, an utterance by a female speaker was used as \( s_1(t) \), and a speech sample uttered by a male speaker was used as \( s_2(t) \). Both \( s_1(t) \) and \( s_2(t) \) were sampled at 8 [kHz], and then these were mixed on a computer so that \( \theta_1 = 12.2 \) [degree] and \( \theta_2 = 58.2 \) [degree], where \( d \) and \( e \) were respectively assumed that \( d = 0.2 \) [m] and \( 340 \) [m]. The constraint was set to \( N = 1 \), \( \phi_1 = \theta_1 \), and \( h_1 = 1 \). For the adaptive learning algorithm, the step-size parameters were set to \( \mu = 0.5 \) for the LMS and \( \beta = 0.9997 \) for the SMI. Other parameters were chosen as \( P = 62 \), \( \tau_0 = 31 \), and \( Q = 62 \).

For the objective measurement, in terms of the instantaneous SNR was considered. The instantaneous SNR of the zero-mean variable is defined by

\[
SNR(t) = 10 \log_{10} \frac{\sum_{l=0}^{L-1} |s(t + l - \tau_0)|^2}{\sum_{l=0}^{L-1} |s(t + l - \tau_0) - \hat{s}(t + l)|^2},
\]

(30)

where \( s(t) \) and \( \hat{s}(t) \) are the target and retrieved signals, respectively. \( \tau_0 \) is inserted by taking the constant delay in (22) into account. In this simulation, \( L \) was set to 512.

Figures 3 and 4 show simulation results of the proposed DCMP criterion. As shown in Figs.3(d) and 4, the proposed method reconstructed the target speech efficiently. Furthermore, in Figs.3(e) and (f), the LMS showed relatively slow convergence on the optimum solution, while the SMI converged faster than the case of the LMS. Hence, it can be said that the proposed DCMP criterion is applicable to broadband signals.

6. Conclusion

A novel algorithm of adaptive beamforming has been presented. In the method, the conventional DCMP criterion for narrowband signals has been extended to broadband signal. The method discriminate between target and undesirable signals by summing up the FIR filtered input at each channel. Since the proposed method requires no time-delay systems as pre-processing of sensor array, the algorithm can be constructed by a relatively small hardware amount of the system.

7. References