ROBUST ADAPTIVE BEAMFORMING ALGORITHM USING INSTANTANEOUS DIRECTION OF ARRIVAL WITH ENHANCED NOISE SUPPRESSION CAPABILITY

Byung-Jun Yoon[1], Ivan Tashev[2], and Alex Acero[2]

[1] Dept. of Electrical Engineering, California Institute of Technology
Pasadena, CA 91125, USA, bjyoon@caltech.edu
[2] Microsoft Research, One Microsoft Way, Redmond, WA 98052, USA
{ivantash, alexac}@microsoft.com

ABSTRACT

In this paper, we propose a novel adaptive beamforming algorithm with enhanced noise suppression capability. The proposed algorithm incorporates the sound-source presence probability into the adaptive blocking matrix, which is estimated based on the instantaneous direction of arrival of the input signals and voice activity detection. The proposed algorithm guarantees robustness to steering vector errors without imposing ad hoc constraints on the adaptive filter coefficients. It can provide good suppression performance for both directional interference signals as well as isotropic ambient noise. For in-car environment the proposed beamformer shows SNR improvement up to 12 dB without using an additional noise suppressor.

Index Terms — Array Signal Processing, Adaptive Arrays, Speech Enhancement, Generalized Sidelobe Canceller (GSC), Instantaneous Direction of Arrival (IDOA)

1. INTRODUCTION

Microphone arrays have been widely studied because of their effectiveness in enhancing the quality of the captured audio signal. The use of multiple spatially distributed microphones allows performing spatial filtering along with conventional temporal filtering, which can better reject the interference signals, resulting in an overall improvement of the captured sound quality.

Various microphone array processing algorithms have been proposed [1], among which the generalized sidelobe canceller (GSC) architecture has been especially popular. The GSC is an adaptive beamformer that keeps track of the characteristics of the interfering signal, leading to a high interference rejection performance. However, if the actual direction of arrival (DOA) of the target signal is different from the expected DOA, a considerable portion of the target signal will leak into the interference canceller, which results in target signal cancellation.

In order to avoid target signal cancellation, a number of methods have been proposed to make the whole system more robust to steering vector errors [2, 3, 4]. The robust GSC (RGSC) proposed in [2] uses an adaptive blocking matrix (ABM) with coefficient-constrained adaptive filters, which prevents the target signal from leaking into the adaptive interference canceller (AIC). In addition to this, the AIC uses norm-constrained adaptive filters that can further improve the robustness against target signal cancellation. This adaptive beamformer is robust to DOA errors without significant degradation in the interference rejection capability. In [3], a similar RGSC has been implemented in frequency-domain, with a comparable performance at a lower computational cost.

Although these RGSCs [2,3] are good at rejecting directional interference signals (such as "jammer" signals), their noise suppression capability is not satisfactory under the conditions of practically isotropic ambient noise, therefore these algorithms are not quite suitable for noisy environments such as the automotive environment.

In this paper, we propose a new adaptive beamforming algorithm that is robust to steering vector errors and that can provide good rejection performance for both directional interference signals and isotropic ambient noise. The proposed algorithm estimates the presence probability of the target signal based on the instantaneous direction of arrival (IDOA) [5] and per frequency bin voice activity detection, which is incorporated into the ABM.

2. ROBUST GENERALIZED SIDELOBE CANCELLER

The structure of the robust GSC proposed in [2] is illustrated in Fig. 1, for a four-element microphone array[1]. Initially, the four microphone inputs $x_i(n)$ ($i = 1,...,4$) go through the fixed beamformer (FBF) that directs the beam

---

[1] For simplicity, we do not show the delays that are needed to guarantee the causality of the system.
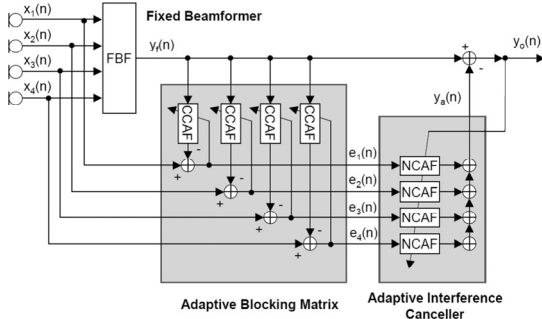
Figure 1. The structure of the robust generalized sidelobe canceller proposed by Hoshuyama et al. [2].

towards the expected DOA. The beamformer output $y_f(n)$ contains the enhanced signal originating from the pointed direction, which is used as a reference by the ABM. The ABM adaptively subtracts the signal of interest, represented by the reference signal $y_f(n)$, from each channel input $x_i(n)$, and provides the interference signals $e_i(n)$. In order to suppress only those signals that originate from a specific tracking region, the adaptive filter coefficients are constrained within predefined boundaries [2]. These boundaries are specified based on the maximum allowed deviation between the expected DOA and the actual DOA.

The interference signals $e_i(n)$, obtained from the ABM, are passed to the AIC, which adaptively removes the signal components that are correlated to the interference signals from the beamformer output $y_f(n)$. The norm of the filter coefficients in the AIC is constrained to prevent them from growing excessively large. This minimizes undesirable target signal cancellation, when the target signal leaks into the AIC [2], further improving the robustness of the system.

Ideally, the beamformer output $y_f(n)$ should consist of the target signal, while the ABM outputs contain only the interference part in the original inputs. This holds more of less true when the interference is a point source whose direction is sufficiently separated from the direction of the target signal. This assumption breaks down under conditions of isotropic ambient noise, where the beamformer output contains a considerable amount of noise. In such cases the ABM tends to subtract the noise from the input signal, while leaving still a significant portion of the target signal behind. Therefore, $e_i(n)$ will have less noise and more from the target signal, which reduces the noise suppression capability of the AIC and leads to target-signal cancellation, degrading the quality of the final output of the RGSC.

3. IDOA-BASED ROBUST GSC

The structure of the proposed adaptive beamformer is shown in Fig. 2 for a four-element microphone array. We first convert the time-domain signal $x_i(n)$ to frequency-
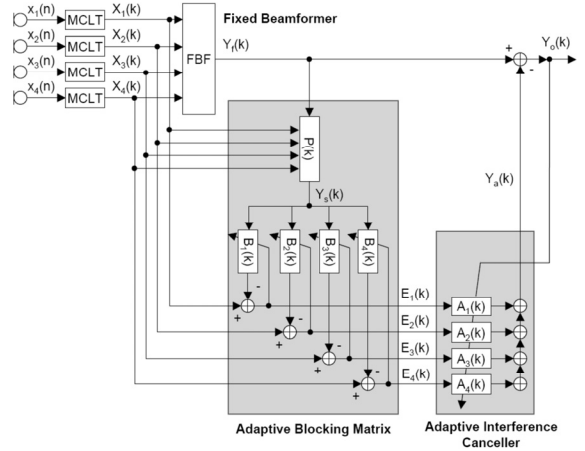

Figure 2. The structure of the proposed IDOA-based robust adaptive beamformer.

domain using modulated complex lapped transform (MCLT) that can give good separation between frequency-bands in an efficient manner [6]. We denote the MCLT of the i-th input signal $x_i(n)$ as $X_i^\ell(k)$, where k is the frequency bin and ℓ is the index of the time-frame. The beamformer output is a weighted sum of the input signals

$$Y_f^\ell(k) = \sum_i W_i(k)X_i^\ell(k). \tag{1}$$

3.1. Adaptive Blocking Matrix

The motivation for passing the beamformer output $Y_f^\ell(k)$ to the ABM is to use it as a reference signal that closely resembles the desired target signal. The ABM should subtract only the portions in the input signals that are correlated to the "desired signal" in the beamformer output. Based on this reasoning, we first estimate the desired signal by taking the product of $Y_f^\ell(k)$ and the probability $P^\ell(k)$ that $Y_f^\ell(k)$ contains the desired signal originating from the target region

$$Y_s^\ell(k) = P^\ell(k)Y_f^\ell(k). \tag{2}$$

The ABM uses the estimated desired signal $Y_s^\ell(k)$ as the reference, instead of the beamformer output $Y_f^\ell(k)$, and adaptively subtracts the components that are correlated to $Y_s^\ell(k)$ from the input signals

$$E_i^\ell(k) = X_i^\ell(k) - B_i^\ell(k)Y_s^\ell(k). \tag{3}$$

After the subtraction, the ABM outputs $E_i^\ell(k)$ contain the interference signals that come from undesirable directions and the background noise originating from the target region.

The adaptive filter coefficients $B_i^\ell(k)$ are updated using the normalized least mean squares (NLMS) algorithm

$$B_i^{\ell+1}(k) = B_i^\ell(k) + \mu_b^\ell(k)Y_s^{\ell\,*}(k)E_i^\ell(k), \tag{4}$$
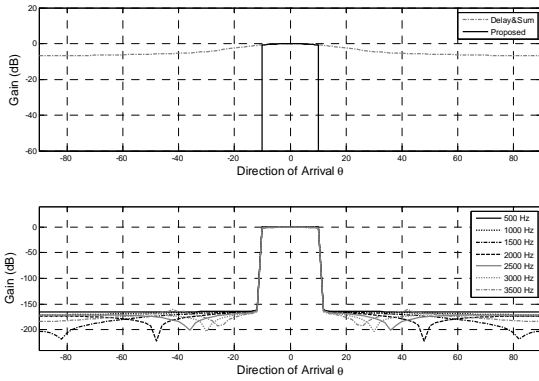
Figure 3. Simulation results for spatial suppression of the proposed robust adaptive beamformer, (Top) for a bandpass signal (300–3400 Hz), and (Bottom) for single tone signals.

where the normalized step size $\mu_b^\ell(k)$ at the ℓ-th frame is computed as follows:

$$\mu_b^\ell(k) = \mu_b \left[\varepsilon_b + S_f^\ell(k)\right]^{-1}. \tag{5}$$

$\mu_b(k)$ is a fixed step size parameter and ε_b is a small number that is used to prevent $\mu_b^\ell(k)$ from becoming too large. $S_f^\ell(k)$ is the power estimate of the beamformer output in the k-th frequency bin at the ℓ-th frame

$$S_f^{\ell+1}(k) = \lambda_b S_f^\ell(k) + (1-\lambda_b)\left|Y_f^\ell(k)\right|^2. \tag{6}$$

The parameter λ_b can be used to control the update speed.

3.2. Estimating the Sound-Source Presence Probability

$P^\ell(k)$ is defined as the probability that the signal of interest originating from the target region is presented in the k-th frequency bin of the beamformer output $Y_f^\ell(k)$ at time-frame ℓ. This probability can be computed as follows:

$$P^\ell(k) = P_s^\ell(k)P_v^\ell(k). \tag{7}$$

The first term, $P_s^\ell(k) = \int_{\theta_T - \delta_\theta}^{\theta_T + \delta_\theta} p_k^\ell(\theta)d\theta$, is the probability that the signal component in k-th frequency bin originates from inside the target tracking region $[\theta_T - \delta_\theta, \theta_T + \delta_\theta]$, where θ_T is the expected DOA of the target signal and δ_θ is the maximum allowable DOA error. We estimate $p_k^\ell(\theta)$ using the method proposed in [5]. With the spatial probability $P_s^\ell(k)$ we have a direct control over the maximum allowable DOA error. Unlike the previous RGSCs [2, 3], we do not have to impose ad hoc constraints on the ABM filter coefficients, which indirectly affect the maximum target direction error. The speech presence probability $P_v^\ell(k)$ in the k-th frequency bin, which is the second term in the right-

hand side of (7), can be computed using a voice activity detector (VAD). In our implementation, we used the VAD based on [7], modified to give noise robust performance.

3.3. Adaptive Interference Canceller

The AIC should remove the signal components in the beamformer output $Y_f^\ell(k)$ that are correlated to the interference signals. First the AIC computes the overall interference signal by taking the weighted sum of $E_i^\ell(k)$ using the adaptive filter coefficients $A_i^\ell(k)$

$$Y_a^\ell(k) = \sum_i A_i^\ell(k)E_i^\ell(k). \tag{8}$$

Then this signal $Y_a^\ell(k)$ is subtracted from $Y_f^\ell(k)$ to eliminate the remaining interference and noise components in the beamformer output

$$Y_o^\ell(k) = Y_f^\ell(k) - Y_a^\ell(k). \tag{9}$$

The final output $y_o(n)$ can be obtained by taking the inverse-MCLT (IMCLT) of $Y_o^\ell(k)$. As in the ABM, the filter coefficients in the AIC are updated using the NLMS algorithm:

$$A_i^{\ell+1}(k) = A_i^\ell(k) + \mu_a^\ell(k)E_i^{\ell\,*}(k)Y_o^\ell(k) \tag{10}$$

The normalized step size $\mu_a^\ell(k)$ is defined as

$$\mu_a^\ell(k) = \mu_a\left[\varepsilon_a + S_a^\ell(k)\right]^{-1}, \tag{11}$$

where $S_a^\ell(k)$ is the power estimate of the interference signals $E_i^\ell(k)$

$$S_a^{\ell+1}(k) = \lambda_a S_a^\ell(k) + (1-\lambda_a)\sum_i\left|E_i^\ell(k)\right|^2. \tag{12}$$

We can also impose norm-constraints on the AIC coefficients as the other RGSCs [2,3,4], but as long as the probability estimator works properly, these constraints are usually not necessary.

4. EXPERIMENTAL RESULTS

In order to evaluate the proposed adaptive beamforming algorithm, we carried out several experiments using a four-element linear microphone array of length 160 mm, where the distances between the adjacent microphones, from left to right, were 45 mm, 70 mm, and 45 mm. Multichannel USB board and a Tablet PC were used to capture the input signals at 16 kHz/16 bits. All processing was done in MCLT domain with 320 samples frames, 50% overlapped.

4.1. Spatial suppression

We computed the spatial suppression of the adaptive beamformer by generating synthetic signals with different inci-
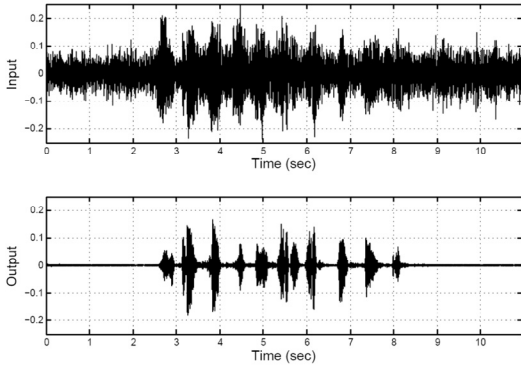
Figure 4. (Top) Original input signal recorded in a car while driving. (Bottom) The output of the proposed beamformer.

dent angles. As an input, we used white Gaussian noise passed through a band-pass filter (300–3400 Hz). The incident angle of the input signal was varied from –90° to 90°, while the expected target DOA was set to 0°. In this experiment we used only the spatial probability $P_s^\ell(k)$ to determine whether the input signal originated from the target region or not. The maximum allowed DOA error was set to 10°. The parameters for the ABM and the AIC were $\lambda_b = \lambda_a = 0.05$, $\varepsilon_b = \varepsilon_a = 0.001$, $\mu_b = 1.0$, and $\mu_a = 0.3$. The gain of the proposed algorithm was computed from the ratio between the output power and the input power after the convergence of the adaptive filters (less than 0.5 sec).

The spatial suppression for band-pass signal is shown in Fig. 3 (Top), together with directivity of a simple delay&sum beamformer. The proposed algorithm is robust to target direction errors and for signals that originate from the target region ($-10° \le \theta \le 10°$ in this case), there is practically no suppression (–0.7 dB at $\theta = \pm10°$). The signals coming from undesirable directions are almost completely suppressed. Fig. 3 (Bottom) shows the suppression for several single-tone signals at different frequencies. The frequency-dependency of the proposed algorithm is negligible, providing good performance at all frequencies. We should note here a non-correlated instrumental noise in the input channels will reduce the suppression capabilities of the adaptive beamformer.

4.2. Experiments with In-Car Recordings

We also tested the proposed beamformer using real signals recorded inside a car. The microphone array was installed on the dashboard between the driver and the passenger. The distance between the center of the microphone array and the passenger was about 70 cm, and the horizontal incident angle was around 22°. The recorded signals contain ambient noise with significant amount of non-stationary components from passing cars, engine noise, whirling wind, and so on. The parameters for the adaptive filters were $\lambda_b = \lambda_a = 0.05$, $\varepsilon_b = \varepsilon_a = 0.001$, $\mu_b = 1.0$, and $\mu_a = 0.15$. Fig.4 shows one

Table 1. SNR improvement for in-car conditions.

Driving condition	Recorded speech	Input SNR	Output SNR	SNR Improvement
Local	Male	12.0	24.0	12.0
Local	Female	12.0	21.0	9.0
Freeway	Male	4.3	14.2	9.9
Freeway	Female	4.4	13.2	8.8

of the input channels (Top) and the output of the proposed beamformer (Bottom). All numbers are in dB. The improvement in SNR is ~10 dB with minimal distortions and low level musical noise. Table 1 shows the SNR improvements attained by the proposed beamformer for several input signals recorded under different conditions. Note that all the SNR improvements have been measured in the telephone-band (300–3400 Hz). The proposed beamformer achieves good noise suppression results without using additional noise suppression techniques.

5. CONCLUDING REMARKS

In this paper, we have proposed a novel adaptive beamforming algorithm based on the RGSC architecture, which incorporates the sound source presence probability estimated using a IDOA-based spatial probability estimator and a voice activity detector. The proposed algorithm is robust to steering vector errors, and provides good noise suppression. It was evaluated in a real automotive environment with high level non-stationary noise, where it was able to eliminate most of the background noise in the original recording with minimal distortion and negligible musical noise.

6. REFERENCES

[1] M. Brandstein and D. Ward, *Microphone Arrays*, Springer Verlag, 2001.

[2] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters", *IEEE Trans. on Signal Processing*, vol. 47, pp. 2677-2684, Oct. 1999.

[3] W. Herbordt and W. Kellermann, "Computationally efficient frequency-domain robust generalized sidelobe canceller", *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Darmstadt, Sep. 2001.

[4] H. Cox, R. M. Zeskind, and M. M. Owen, "Robust adaptive beamforming", *IEEE Trans. on Acoustics, Speech, Signal Processing*, vol. 35, pp. 1365-1379, Dec. 1986.

[5] I. Tashev and A. Acero, "Microphone array post-processing using instantaneous direction of arrival", *Proc. International Workshop on Acoustic Echo and Noise Control (IWAENC)*, Paris, Sep. 2006.

[6] H. S. Malvar, "A modulated complex lapped transform and its applications to audio processing", *Proc. IEEE Int. Conf. on Acoustics, Speech, and Signal Processing (ICASSP)*, Phoenix, Mar. 1999.

[7] J. Sohn, N. S. Kim, and W. Sung, "A statistical model-based voice activity detection", *Signal Processing Letters*, vol. 6, pp. 1-3, Jan. 1999.