A SIMPLE MODIFICATION TO FACILITATE ROBUST GENERALIZED SIDELOBE CANCELLER FOR HEARING AIDS

Meng Guo and Jan Mark de Haan and Jesper Jensen

Oticon A/S
Kongebakken 9
DK-2765 Smørum
Denmark
emails: {guo,jmh,jsj}@oticon.dk

ABSTRACT

This work focuses on an adaptive beamformer in a hearing aid application using a generalized sidelobe canceller structure (GSC). In this application, the constraint and blocking matrices in the GSC structure are specifically designed using an estimate of the transfer functions between the target source and the microphones to ensure optimal beamformer performance. We show that, in practice, the GSC always—unintentionally—attenuates the target sound in a special but realistic situation where all signals, including the target and noise signals, originate from the look direction reflected by the look vector. This happens because, in practice, the blocking matrix in the GSC structure is non-ideal. We introduce a *simple* modification to the GSC structure, which solves the problem of undesired target signal attenuation in situations where all signals originate from the look direction. Furthermore, this modification can also prevent desired signals, originating from positions spatially close to the look direction, to be removed. We also show that the solution has no impact on other acoustic situations.

Index Terms— Generalized sidelobe canceller, look vector, target-cancelling beamformer.

1. INTRODUCTION

In hearing aids, a microphone array beamformer is often used for spatially attenuating background noise sources [1]. Many beamformer variants can be found in literature, see, e.g., [2] and the references therein. The minimum variance distortionless response (MVDR) beamformer is widely used in microphone array signal processing, see some recent examples in [3–6]. Ideally the MVDR beamformer keeps the signals from the target direction (also referred to as the look direction) unchanged, while attenuating sound signals from other directions maximally. The generalized sidelobe canceller (GSC) structure is an equivalent representation of the MVDR beamformer offering computational and numerical advantages over a direct implementation in its original form, see, e.g., [7, 8]. In this work, we focus on the GSC structure in a hearing aid application.

Fig. 1 illustrates the GSC structure. All signals are represented in the frequency domain for convenience. The target source signal is denoted by $s(k,n)$, where k is the frequency index and n is the time index, $d_m(k)$ is the transfer function from $s(k,n)$ to the mth microphone, where $m = 1, ..., M$, and the microphone signals are denoted by $y_m(k,n)$. For convenience, we assume these transfer functions to be time-invariant. Furthermore, $\mathbf{c}(k) \in \mathbb{C}^{M\times 1}$ denotes the time-invariant constraint vector, which is also referred

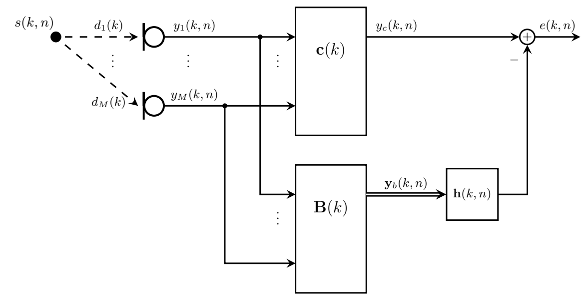

Fig. 1. An overview of the GSC structure in the frequency domain.

to as an all-pass beamformer, whereas $\mathbf{B}(k) \in \mathbb{C}^{M\times(M-1)}$ denotes the blocking (or target-cancelling) beamformer. The optimal filter vector $\mathbf{h}(k,n) \in \mathbb{C}^{(M-1)\times 1}$ is obtained by minimizing the mean square error (MSE) of the GSC output signal $e(k,n)$. Ideally, the all-pass beamformer $\mathbf{c}(k)$ does not modify the target signal from the look direction. The target-cancelling beamformer $\mathbf{B}(k)$ is orthogonal to $\mathbf{c}(k)$, and it has nulls in the look direction, hence, ideally it removes the target source signal completely.

For simplicity, we focus on the case with only two microphones, i.e., $M = 2$. However, the theory and results obtained can be easily adapted for cases where $M > 2$. As a result of choosing $M = 2$, the matrix $\mathbf{B}(k)$ becomes a vector $\mathbf{b}(k)$, its output signal vector $\mathbf{y}_b(k,n)$ is a scalar $y_b(k,n)$, and the optimal filter vector $\mathbf{h}(k,n)$ is a scaling factor $h(k,n)$.

It is well-known that the MVDR beamformer can, despite the distortionless response constraint, cancel the desired signal from the look direction. This would, e.g., be the case in a reverberant room, when reflections of the desired target signal pass through the target-cancelling beamformer, and its output signal $y_b(k,n)$ is thereby correlated with the target signal. Target cancellation can also occur due to look vector estimation errors [9, 10]. Some sophisticated solutions to this problem exist, such as introducing an adaptive target-cancelling beamformer $\mathbf{B}(k,n)$ [11], or taking the probability of look vector errors into account when designing the beamformer [12], and the suggestion of a more accurate look vector estimation [13].

In this paper, we propose a simple solution to a specific instance of this problem which occurs in some realistic situations. As we explain in more detail below, the undesired target signal cancellation is unavoidable in practical applications, due to non-ideal target-cancelling beamformers $\mathbf{b}(k)$.

In a hearing aid application, the microphone array is typically placed closely to the ear of the hearing aid user to ensure that the array picks up the most realistic sound signals for a natural sound perception. Therefore, the transfer functions $d_m(k)$ vary for different hearing aid users. We define the look vector $\mathbf{d}(k)$ as $\mathbf{d}(k) = [d_1(k), ..., d_M(k)]^T$.

In practical applications, the look vector $\mathbf{d}(k)$ is unknown, and it must be estimated. This is typically done in a calibration procedure in a sound studio with a hearing aid mounted on a head-and-torso-simulator. Furthermore, the beamformer coefficients of $\mathbf{c}(k)$ and $\mathbf{b}(k)$ are constructed based on the look vector estimate $\hat{\mathbf{d}}(k)$. This is further explained in Sec. 2.

As a result of using the look vector estimate $\hat{\mathbf{d}}(k)$ rather than $\mathbf{d}(k)$, the target-cancelling beamformer $\mathbf{b}(k)$ does not have a perfect null in the look direction, it has a finite attenuation of typically $10-30$ dB as observed in our measurements. This phenomenon allows the GSC to—unintentionally—attenuate the target source signal while minimizing the GSC output signal $e(k,n)$, as we will describe more closely in Sec. 3. A solution to this problem will be introduced in Sec. 4 and it is evaluated in Sec. 5.

In this paper, column vectors and matrices are emphasized using lower and upper letters in bold, respectively. Transposition, Hermitian transposition and complex conjugation are denoted by the superscripts T, H and $*$, respectively.

2. ALL-PASS AND TARGET-CANCELLING BEAMFORMERS

In this section, we describe the design of the beamformers $\mathbf{c}(k)$ and $\mathbf{b}(k)$. Beam patterns of the two beamformers will be depicted in free field and in a realistic (measured) acoustic field. For convenience, we drop the frequency index k in the following; however, all processing is performed independently for each frequency index.

2.1. Beam Patterns in Free Field

In free field conditions, the look vector $\mathbf{d}$ can be easily determined. We assume that the hearing aid user faces the sound source, and this direction (0 degrees) is defined as the look direction. The target sound and the two microphones are located in the horizontal plane. Using a virtual reference microphone, i.e., $d_{ref} = 1$, located exactly between the physical microphones, the look vector becomes

$$\mathbf{d}_0 = \left[e^{-j\omega\frac{T_d}{2}}, e^{j\omega\frac{T_d}{2}}\right]^T, \tag{1}$$

where $\omega = 2\pi f$, and $T_d = d/c$, where f is the frequency, d is the distance between the two microphones, and c represents the sound speed of $c \approx 340$ m/s. Furthermore, we define a unit-norm version $\mathbf{d}$ of $\mathbf{d}_0$, i.e.,

$$\mathbf{d} = \frac{\mathbf{d}_0}{\|\mathbf{d}_0\|}. \tag{2}$$

The all-pass beamformer $\mathbf{c}$ and the target-cancelling beamformer $\mathbf{b}$ are given by definition

$$\mathbf{c}^H\mathbf{d} = 1 \ \wedge \ \mathbf{b}^H\mathbf{d} = 0. \tag{3}$$

Hence,

$$\mathbf{c} = \mathbf{d}, \tag{4}$$

$$\mathbf{b} = [d_2, -d_1]^H. \tag{5}$$

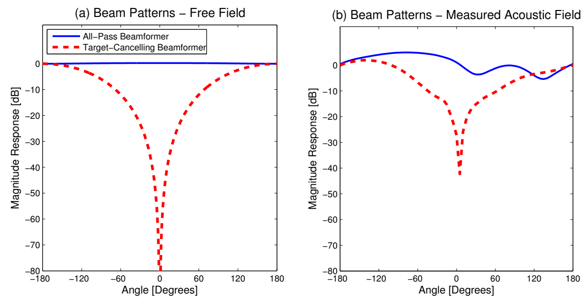

Fig. 2. Beam patterns when the look direction is 0 degrees. (a) Free field. (b) A measured acoustic field.

Inserting (2) in (4) and (5) will result in the beamformer coefficients of these two beamformers. Fig. 2(a) illustrates the beam patterns for an example frequency $f = 1$ kHz of a microphone array with a microphone distance $d = 13$ mm. As expected, the all-pass beamformer $\mathbf{c}$ has unit response in the look direction (0 degrees), whereas the target-cancelling beamformer $\mathbf{b}$ has a perfect null in this direction (Although we can only observe that the magnitude is below -80 dB).

2.2. Beam Patterns in Measured Field

In practice, only an estimate of $\mathbf{d}$ is available. Therefore, we need to derive the beamformer coefficients from the look vector estimate $\hat{\mathbf{d}}$. Hence, equations (4) and (5) become

$$\mathbf{c} = \hat{\mathbf{d}}, \tag{6}$$

$$\mathbf{b} = [\hat{d}_2, -\hat{d}_1]^H. \tag{7}$$

To estimate $\hat{\mathbf{d}}$, we attached a hearing aid to a head-and-torso-simulator in a sound studio. A white noise target signal $s(n)$ was played, impinging from the look direction (0 degrees). The resulting microphone signal vector $\mathbf{y}(n) = [y_1(n), ..., y_M(n)]^T$ is given by

$$\mathbf{y}(n) = s(n)\mathbf{d}. \tag{8}$$

The microphone signal covariance matrix $\mathbf{R}_{yy} = E\left[\mathbf{y}(n)\mathbf{y}^H(n)\right]$, where $E[\cdot]$ is the statistical expectation operator, can be estimated as

$$\hat{\mathbf{R}}_{yy} = \frac{1}{N}\sum_{n=1}^{N}\mathbf{y}(n)\mathbf{y}^H(n), \tag{9}$$

where N determines the duration of the white noise calibration signal $s(n)$. From (9), the look vector estimate $\hat{\mathbf{d}}$ can be found as the eigenvector corresponding to the largest eigenvalue of the covariance matrix estimate $\hat{\mathbf{R}}_{yy}$, where this eigenvector is further normalized to have unit-norm.

Fig. 2(b) illustrates the beam patterns for an example frequency $f = 1$ kHz in a real acoustic field. We observe that the all-pass beamformer only approximates a unity response; more important, the target-cancelling beamformer does not have a perfect null, but it has an attenuation of approximately 35 dB. Increasing the value of N leads to a larger attenuation. However, in real applications, only a finite value of this attenuation can be realized, rather than the theoretically desired response of $-\infty$ dB when $\lim_{N\to\infty}\hat{\mathbf{d}} = \mathbf{d}$.

As we explain in more detail in Sec. 3, the problem arises because the attenuation is finite, i.e., increasing N cannot solve the problem. In other words, we always suffer from the target-cancelling problem whenever $N \neq \infty$ and we only obtain a finite attenuation of the target signal from the look direction.

3. THE TARGET-CANCELLING PROBLEM

This section describes how the GSC minimizes its output signal $e(n)$, and subsequently the target-cancelling problem.

The GSC output signal $e(n)$ is expressed by

$$e(n) = y_c(n) - h(n)y_b(n), \tag{10}$$

as shown in Fig. 1. To ensure that the GSC does not attenuate desired speech signals, the scaling factor $h(n)$ is found during noise-only periods, i.e., when a voice activity detection (VAD) declares noise only. The computation of $h(n)$ is expressed by

$$h_{opt}(n) = \arg\min_{h(n)} E[|e(n)|^2], \text{ when VAD} = 0. \tag{11}$$

The closed-form solution of (11) is

$$h(n) = \frac{E[y_b^*(n)y_c(n)]}{E[y_b^*(n)y_b(n)] + \delta}, \text{ when VAD} = 0, \tag{12}$$

where $\delta > 0$ is a regularization parameter.

In the following, we discuss the target-cancelling problem in a specific but realistic situation, where the target and all noise signals originate from the look direction.

In the ideal situation, the output signal $y_c(n)$ of the all-pass beamformer $\mathbf{c}$ contains the mixture of the target and the noise signals due to the unity response in the look direction. The output signal $y_b(n)$ should ideally be zero due to a perfect null in the target-cancelling beamformer $\mathbf{b}$ at the look direction, as illustrated in Fig. 2(a). By analyzing (12), we obtain $h(n) = 0$ since $\delta > 0$; hence, we obtain $e(n) = y_c(n)$, i.e., all signals pass through the GSC structure unmodified. This result is desired in this situation, since all signals originate from the look direction.

However, in practice, the target-cancelling beamformer $\mathbf{b}$ does not have a perfect null as illustrated in Fig. 2(b); it has a relatively large but finite attenuation in the look direction, such as 40 dB. Analyzing again (12), we observe that the numerator $E[y_b^*(n)y_c(n)]$ now has a nonzero value, and the first part of the denominator $E[y_b^*(n)y_b(n)]$ is also non-zero and numerically less than the numerator. When the regularization parameter δ has a comparably smaller numerical value, the resulting scaling factor $h(n)$ would be $h(n) \neq 0$, which is undesirable.

Fig. 3 shows a simulation of $h(n)$, estimated according to (12), in 16 frequency channels in the acoustic field used in Fig. 2(b). The expectation operator in (12) is replaced by a first-order IIR smoothing filter with a time constant $\tau = 200$ ms, where $e^{-t/\tau} \approx 0.37$, and t denotes the time. We clearly observe that $h(n) \neq 0$, and $|h(n)| \approx 30$ dB in some frequency channels; the target-cancelling problem thereby arises, as the target-cancelling beamformer attenuates the target signal by approximately 30 dB at these frequencies.

Furthermore, Fig. 4 shows the transfer function of the GSC for signals from the look direction. Ideally, it should be 0 dB for all frequencies, but due to the non-ideal target-cancelling beamformer $\mathbf{b}$ and the update procedure of $h(n)$ in (12), the obtained response is far from the desired. An attenuation of more than 30 dB is observed at some frequencies.

In fact, the response in Fig. 4 can somehow be considered as an exaggerated example to demonstrate the problem, since all signals originate from the look direction. However, the target-cancelling problem would also have influence, although reduced, in other situations, e.g., with a dominating target signal from the look direction, and low-level noise signals are coming from other directions.

Additionally, if the target source is located just off the look direction, e.g., 5 degrees to one side because the hearing aid user is

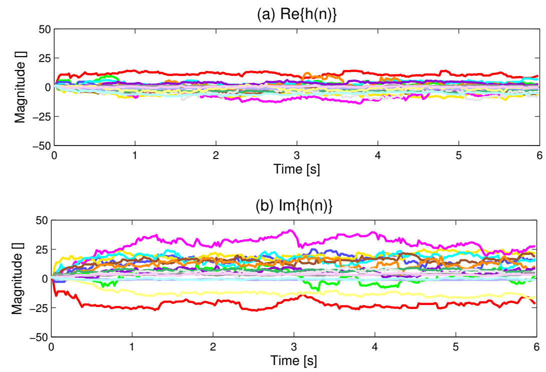

Fig. 3. Numerical values of $h(n)$ in 16 example frequency channels with the ordinary GSC structure. (a) Real parts. (b) Imaginary parts.

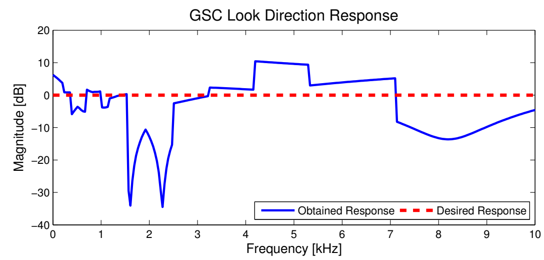

Fig. 4. The GSC magnitude response of the look direction. The resulting response is not as expected due to $h(n) \neq 0$.

not facing directly to the sound source, then this source signal would pass through the target-cancelling beamformer with a finite attenuation, both in the ideal or non-ideal situations as illustrated in Fig. 2. The GSC structure will somewhat remove this signal even though it is considered to be the target signal. The solution described in Sec. 4 can also be used to avoid this undesired cancellation of signals spatially close to the look direction.

4. MODIFIED GSC SCALING FACTOR UPDATE

To resolve the target-cancelling problem, we introduce a modification to the scaling factor update in (12) for each frequency channel. The simplicity of this solution makes it attractive in hearing aids with only limited processing power.

Recall that the problem in the specific case where all signal sources are located in the look direction was caused by a non-ideal target-cancelling beamformer $\mathbf{b}(n)$; as a consequence, the denominator gets smaller than the numerator in (12). A fixed regularization parameter δ cannot solve this problem, since the target source level affects the numerical values of the numerator and the denominator.

To solve this problem, we introduce a dependency of the estimation of $h(n)$ on the difference $\Delta(n)$ between the energy of the beamformer output signals $y_c(n)$ and $y_b(n)$, expressed by

$$\Delta(n) = \frac{\sum_{l=0}^{L-1} |y_c(n-l)|^2}{\sum_{l=0}^{L-1} |y_b(n-l)|^2}, \tag{13}$$

where L is the number of data samples used to compute $\Delta(n)$.

The difference $\Delta(n)$ is largest, when all signal sources are located in the look direction. This would be the case for either ideal or non-ideal target-cancelling beamformer $\mathbf{b}$, since the target-cancelling beamformer has a null (even it was non-ideal) in the

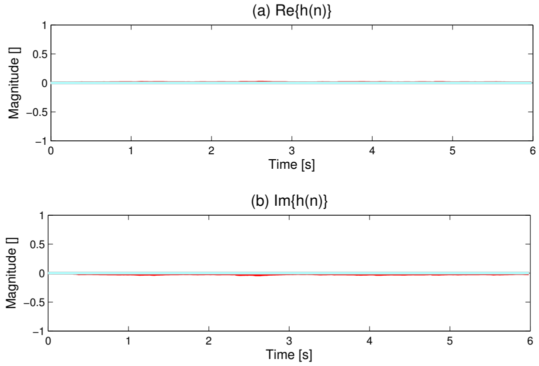

Fig. 5. Numerical values of $h(n)$ in 16 example frequency channels with the modified GSC structure. (a) Real parts. (b) Imaginary parts.

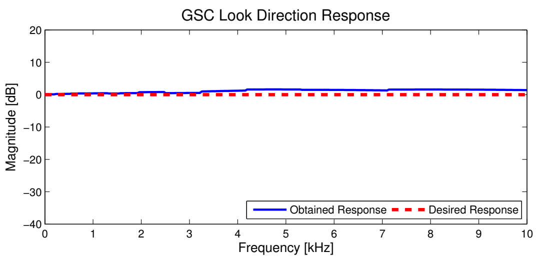

Fig. 6. The magnitude response of the modified GSC structure for signals impinging from the look direction.

look-direction, see also the examples in Fig. 2. Therefore, we monitor the difference $\Delta(n)$ to control the estimation of the scaling factor. A modified scaling factor $h_{mod}(n)$ is thereby introduced, and it is defined as

$$h_{mod}(n) = \begin{cases} h(n) & \text{for } \Delta(n) \leq \eta, \\ 0 & \text{otherwise.} \end{cases} \quad (14)$$

The threshold value η is determined by the difference between the magnitude responses of the all-pass beamformer $\mathbf{c}$ and the target-cancelling beamformer $\mathbf{b}$ in the look direction. In the example shown in Fig. 2(b), an appropriate η-value would for instance be $\eta = 30$ dB.

5. VERIFICATIONS

In this section, we verify the proposed GSC modification. Again, we consider the real acoustic situation in Fig. 2(b), which suffered from the undesired target-cancelling problem as shown in Fig. 4. We use the modified estimate $h_{mod}(n)$ defined in (14), and the η-values across frequencies are set to values between 5 and 30 dB; the exact value in each frequency is found based on the look vector estimate $\hat{\mathbf{d}}$; a value of 30 dB is used in the case shown in Fig. 2(b).

Fig. 5 shows the resulting $h_{mod}(n)$ values obtained in a computer simulation. We observe that the obtained $h_{mod}(n)$ values are all close to zero as expected. Furthermore, Fig. 6 illustrates the magnitude response of the GSC for signals impinging from the look direction, and we observe that the magnitude response is very close to the ideal 0 dB for all frequencies; the target-cancelling problem seen in Fig. 4 is thereby resolved as expected. The look direction response is also successfully verified in a sound studio experiment with a real hearing aid, although the result is not shown in this paper.

Table 1. Mean square errors [dB] of target input/output signals.

	GSC	**Modified GSC**
Look Direction Only	-5.5	-12.5
Car	-6.1	-6.1
Lecture	-6.4	-6.4
Meeting	-16.2	-16.2
Party	-10.1	-10.1
Restaurant	-9.0	-9.0

Moreover, we carried out simulation experiments to verify that while the modification resolves the problem in the specific case where all source signals impinge from the look direction (0 degrees), it has no negative influence on other situations. Five additional sound environments were considered, which are good representatives of a hearing aid user's (HAU) everyday life:

Car The HAU wants to listen to one person on the passenger seat while driving a car.

Lecture The HAU wants to listen to one person far away in a very large room with reverberation and ambient noise.

Meeting The HAU wants to listen to two people talking to each other, in a room with strong reflections.

Party The talker is close-by in front of the HAU. Most disturbing people are further away, but one is close-by from the behind.

Restaurant The HAU wants to listen to two people talking to each other, amongst low-level background noise as in a restaurant.

We evaluate the modified GSC performance using the MSE between the input and output target signals. In the simulations, the input signals are created as mixtures of target and noise signals. The GSC is applied on these mixture signals, and the obtained scaling factors $h_{mod}(n)$ are subsequently used to process the input target and noise signals separately to compute the target and noise components in the GSC output signals. Before the MSE calculation, we also compensated for processing delay between the input and output target signals. The MSEs were computed for both the unmodified GSC and the modified GSC, the results are given in Table 1.

Table 1 shows that in the case where all source signals impinge from the look direction, and the input mixture signal contains a speech signal in noise, the GSC has a relatively large MSE (-5.5 dB) compared to the modified GSC (-12.5 dB), indicating that the undesired target signal cancellation took place in the GSC, and the modified GSC resolves the problem, as expected. Otherwise, there is no difference between these two GSC setups in the other five representative sound environments, indicating that the proposed GSC modification does not introduce artifacts in other situations.

6. CONCLUSION

In this paper, we addressed a problem which occurs when using a GSC structure in a hearing aid application. The problem arises due to a non-ideal target-cancelling beamformer. As a consequence, a target signal impinging from the look direction can—unintentionally—be attenuated by as much as 30 dB. To resolve this problem, we monitored the difference between the output signals from the all-pass beamformer and the target-cancelling beamformer to control the GSC update. The greatest advantage of this proposed solution is its simplicity, which is a crucial factor in a hearing aid with only limited computational power. Simulation results also verified that it resolves the target-cancelling problem without introducing other artifacts.

7. REFERENCES

[1] H. Dillon, *Hearing Aids*. Stuttgart, Germany: Thieme, May 2001.

[2] M. Brandstein and D. Ward, *Microphone Arrays: Signal Processing Techniques and Applications*. Berlin, Heidelberg, Germany: Springer, Jun. 2001.

[3] S. Markovich-Golan, S. Gannot, and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer," in *Proc. 2010 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Mar. 2012, pp. 197–200.

[4] R. Heusdens, G. Zhang, R. C. Hendriks, Y. Zeng, and W. B. Kleijn, "Distributed MVDR beamforming for (wireless) microphone networks using message passing," in *Proc. Int. Workshop Acoust. Signal Enhancement*, Sep. 2012.

[5] Z. Jie, W. Qiang, W. Xin-Qun, and Y. Chang-Ming, "Simulation of high-speed train pass-by noise source identification based on MVDR beamforming," in *Proc. 2010 IEEE Int. Conf. Control Autom.*, June. 2013, pp. 254–258.

[6] C. Pan, J. Chen, and J. Benesty, "Performance study of the MVDR beamformer as a function of the source incidence angle," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 1, pp. 67–79, Jan. 2014.

[7] L. J. Griffiths and C. W. Jim, "An alternative approach to lineary contrained adaptive beamforming," *IEEE Trans. Antennas Propag.*, vol. 30, no. 1, pp. 27–34, 1982.

[8] K. M. Buckley, "Broad-band beamforming and the generalized sidelobe canceller," *IEEE Trans. Acoust., Speech, Signal Process.*, vol. 34, no. 5, pp. 1322–1323, 1986.

[9] M. M. Sondhi and G. W. Elko, "Adaptive optimization of microphone arrays under a nonlinear constraint," in *Proc. 1986 IEEE Int. Conf. Acoust., Speech, Signal Process.*, Apr. 1986, pp. 981–984.

[10] G. W. Elko and J. Meyer, "Microphone arrays," *Springer Handbook of Speech Processing*, pp. 1021–1041, Nov. 2007.

[11] O. Hoshuyama, A. Sugiyama, and A. Hirano, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. Signal Process.*, vol. 47, no. 10, pp. 2677–2684, Oct. 1999.

[12] K. L. Bell, Y. Ephraim, and H. L. V. Trees, "A bayesian approach to robust adaptive beamforming," *IEEE Trans. Signal Process.*, vol. 48, no. 2, pp. 386–398, Feb. 2000.

[13] Y. Gu and A. Leshem, "Robust adaptive beamforming based on interference covariance matrix reconstruction and steering vector estimation," *IEEE Trans. Signal Process.*, vol. 60, no. 7, pp. 3881–3885, Apr. 2012.