Robust Adaptation Control for Generalized Sidelobe Canceller with Time-Varying Gaussian Source Model

Jinfu Wang*†, Feiran Yang*†, Jianfeng Guo‡*, and Jun Yang*†
* Institute of Acoustics, Chinese Academy of Sciences, Beijing, China
† University of Chinese Academy of Sciences, Beijing, China
‡ College of Data Science, Taiyuan University of Technology, Taiyuan, China
E-mail: wangjingfu19@mails.ucas.edu.cn, feiran@mail.ioa.ac.cn, guojianfeng@tyut.edu.cn, jyang@mail.ioa.ac.cn

***Abstract*—In the generalized sidelobe canceller (GSC), the adaptation control of the adaptive interference canceller (AIC) should be carefully designed to increase the interference suppression and alleviate the target speech distortion. This paper proposes a robust update rule for the AIC based on the assumption that the output of the beamformer follows a time-varying Gaussian distribution. The coefficients of the AIC and the variance of the target speech are updated in an iterative manner under the maximum likelihood criterion. The proposed method is found to inherently contain a speech detection module that adaptively adjusts the update rate of the AIC. Computer simulations under various conditions confirm the robustness and effectiveness of the proposed control rule over the existing methods.**

***Index Terms*—Generalized sidelobe canceller, speech enhancement, microphone array.**

I. Introduction

The generalized sidelobe canceller (GSC) is an effective adaptive beamformer and has been widely employed for speech enhancement [1], [2]. It consists of three components, namely, the fixed beamformer (FBF), the blocking matrix (BM) and the adaptive interference canceller (AIC). The AIC employs an unconstrained adaptive filter to remove the interference components from the FBF output. The update rule of the AIC should be well designed to avoid the instability of the adaptive filter and the target speech distortion [3]–[11].

The adaptation mode controller (AMC) is very popular for the control of the AIC adaptation [12], [13]. The AMC estimates the signal-to-noise ratio (SNR) with the power ratio between the output of the FBF and the BM, and then the AIC is updated only if the SNR estimate is lower than the predefined threshold. However, the tracking capability of the AMC-AIC is slow, and it is difficult to select a suitable threshold for different scenarios. The speech presence probability (SPP) is utilized to control the adaptation of the AIC [14]. However, unreliable SPP estimate typically occurs in nonstationary noise environments, leading to an unstable adaptation of the AIC. The M-estimate based update rule with variable step size was proposed for a robust update of the AIC [15], which usually combines with the speech detection module (SPP or AMC) to suppress more interference.

The fundamental characteristic of speech signal has also been considered for the robust control of the AIC, where update formulas contain an inherent step-size control mechanism. By exploiting the non-Gaussianity of the target speech, the maximum negentropy (MN) and the maximum kurtosis (MK) criterion were proposed in [16], [17]. The MN-based update rule is more robust to the outliers but with a heavier computational burden than the MK-based one. The statistical independence between different signals has been utilized in [18], [19], and the optimization criterion of the AIC is generalized as minimizing the mutual information. However, these methods have to be implemented in a block-wise way for the robust estimate of statistics, requiring large computational overheads and making them not suitable for an online implementation. For instance, the block-wise based MK-AIC can achieve a satisfactory performance only when the duration of the block is larger than 0.5 s [20].

In this paper, we propose a new update rule for the AIC based on a statistical speech model. Specifically, the time-varying Gaussian source model (TVGSM) is utilized to represent the non-Gaussian characteristic of the target speech signal. With this model, the optimal AIC is derived from the maximum likelihood (ML) perspective, in which the coefficients of the AIC and the variance of the target speech are updated in an iterative way. We show that the online update rules inherently contain an adaptation control module that slows down the adaptation of the AIC during target speech periods, and the proposed ML-AIC can be regarded as the RLS-based algorithm with a variable forgetting factor. Evaluations under varying signal-to-interference ratios (SIRs) and reverberation times demonstrate that the proposed method greatly improves the speech quality and alleviates speech distortion compared to the existing adaptation control methods.

This work was supported by the National Natural Science Foundation of China under Grant 62171438 and 12204345, and IACAS Frontier Exploration Project QYTS202111.

II. SIGNAL MODEL

Consider a microphone array comprising M microphones, which captures a desired speech signal in a noisy reverberant environment. In the short-time Fourier transform (STFT) domain, the microphone signal vector $\mathbf{y}(k,n) = [Y_1(k,n), \cdots, Y_M(k,n)]^T$ can be expressed as [2]

$$\begin{aligned}\mathbf{y}(k,n) &= \mathbf{x}(k,n) + \mathbf{v}(k,n) \\ &= \mathbf{g}_r(k,n)X_r(k,n) + \mathbf{v}(k,n),\end{aligned} \tag{1}$$

where the superscript ^{T} denotes the transpose operator, n is the time index, k is the frequency index, $\mathbf{x}(k,n)$ and $\mathbf{v}(k,n)$ denote desired speech and noise signal vectors, respectively, and $X_r(k,n)$ is the clean speech signal at the reference microphone. The relative transfer function (RTF) vector $\mathbf{g}_r(k,n)$ with respect to the rth microphone is defined as [6]

$$\mathbf{g}_r(k,n) = \left[\frac{G_1(k,n)}{G_r(k,n)}, \cdots, \frac{G_M(k,n)}{G_r(k,n)}\right]^T, \tag{2}$$

where $G_i(k,n)$ is the acoustic transfer function from the source to the ith microphone. Without loss of generality, we assume that the environment is slowly time-varying, and thus $\mathbf{g}_r(k,n)$ can be approximated as $\mathbf{g}_r(k)$. We have already employed the multiplicative transfer function (MTF) approximation [6] in (1). In highly reverberant environments, it is necessary to apply the convolutive transfer function (CTF) approximation [8]. Since the proposed approach can be extended into the CTF-GSC straightforwardly, we limit our discussion to the MTF-GSC in this paper. Note that each frequency bin is treated independently, and we will omit the frequency index k for brevity.

We extract the target speech by the GSC beamformer

$$\hat{X}(n) = \mathbf{w}_{\text{GSC}}^H(n)\mathbf{y}(n), \tag{3}$$

where $\mathbf{w}_{\text{GSC}}(n) = [W_1(n), \cdots, W_M(n)]^T$ is a filter of length M. The conventional GSC minimizes the power of the beamformer output $\hat{X}(n)$ while retaining the target speech undistorted, which can be expressed as [3], [4]

$$\mathbf{w}_{\text{GSC}}(n) = \mathbf{w}_{\text{q}} - \mathbf{B}\mathbf{w}_a(n), \tag{4}$$

where $\mathbf{w}_{\text{q}} \in \mathbb{C}^{M\times 1}, \mathbf{B} \in \mathbb{C}^{M\times(M-1)}$, and $\mathbf{w}_a(n) \in \mathbb{C}^{(M-1)\times 1}$. The fixed beamformer $\mathbf{w}_{\text{q}}$ aims to steer a beam into the target direction and generate a reference $Y_d(n) = \mathbf{w}_{\text{q}}^H\mathbf{y}(n)$ for the desired signal, where a matched filter $\mathbf{w}_{\text{q}} = \mathbf{g}_r/\mathbf{g}_r^H\mathbf{g}_r$ is commonly adopted due to its robustness [6], [13]. The blocking matrix $\mathbf{B}$ is used to completely block the desired signal and to provide ideal noise references $\mathbf{u}(n) = \mathbf{B}\mathbf{y}(n)$ for the interference signals, which can be achieved by spanning the left nullspace of $\mathbf{g}_r$, i.e., $\mathbf{B}^H\mathbf{g}_r = \mathbf{0}$, and then formulating the null towards the direction of the target speech. The AIC typically employs an unconstrained adaptive filter $\mathbf{w}_{\text{a}}(n)$ to eliminate the residual noise in $Y_d(n)$ that is correlated with the noise references $\mathbf{u}(n)$.

In practice, the robust adaptation control for the AIC is indispensable due to the two reasons. Assuming that the BM completely suppresses the target speech signal, the AIC should effectively remove the residual noise correlated with noise references from the output of the FBF. However, during double-talk, i.e., the presence of the target speech signal and noise signals, the AIC may suffer from the instability problem. On the other hand, the BM cannot totally suppress the target speech signal in practice, and the noise references may contain target speech leakage. The continuous update of the AIC in this case would result in the cancelation of the target speech.

In summary, the update of the AIC should be halted or slowed down during the speech periods to ensure the robust adaptation. The AMC-based adaptation control mechanism estimates the SNR through the power ratio between the output of the FBF and the BM, and then halts the update of the AIC if the SNR estimate is lower than the predefined threshold. However, the accuracy of the SNR estimate heavily depends on the design of the FBF and the BM. In addition, the selection of a suitable threshold for different scenarios is not trivial. Other online robust adaptation control mechanisms rely on the more sophisticated speech detection module which often considers the nonstationary characteristic of the speech signal, and they may suffer from performance degradation in nonstationary noise environments.

III. PROPOSED ML-AIC

The TVGSM has been originally utilized to represent the non-Gaussian characteristic of the speech signal in many fields, e.g., source separation, dereverberation and speech enhancement [21]–[23]. In this paper, we adopt the TVGSM as the statistical speech model and reformulate the design criterion of the AIC, where the output of the AIC is ensured to be more like a speech signal in a statistical sense. The optimal AIC is derived from the maximum likelihood perspective. We first present the batch implementation of the proposed update rule, and then we consider an online version that has a better tracking capability and renders a real-time implementation.

A. Batch Algorithm

We suppose that the output of the GSC can be modeled as a zero-mean time-varying Gaussian random variable, i.e., $\hat{X}(n) \sim \mathcal{CN}(0, \lambda(n))$, where $\lambda(n)$ is the variance of $\hat{X}(n)$. Assuming the statistical independence among $\hat{X}(n), n = 1, 2, ...$, the negative log-likelihood function of beamformer output $\hat{X}(n)$ is given by

$$-\mathcal{L}(\mathbf{w}_a, \lambda) = \sum\nolimits_n \ln(\lambda(n)) + \frac{|\hat{X}(n)|^2}{\lambda(n)} + \text{const.} \tag{5}$$

Additionally, we incorporate a regularization term, i.e., the l_2-norm constraint, to the cost function that can penalize the large coefficients of the AIC and further improve the robustness of the GSC [18], [20], and hence we obtain the optimization criterion of the proposed AIC

$$\mathcal{J} = -\mathcal{L}(\mathbf{w}_a, \lambda) + \alpha\mathbf{w}_a^H\mathbf{w}_a, \tag{6}$$

where $\alpha > 0$ is a weight for robustness control. The optimal solution to (6) is derived by minimizing the cost function $\mathcal{J}$, but a closed-form solution for the parameters $\lambda(n)$ and $\mathbf{w}_{\text{a}}$

is not available. We thus choose to iteratively optimize $\mathcal{J}$ by alternately updating $\lambda(n)$ and $\mathbf{w}_{\mathrm{a}}$. By setting the partial derivatives of $\mathcal{J}$ with respect to $\lambda(n)$ and $\mathbf{w}_{\mathrm{a}}$ to zero, the iterative formulas for updating $\lambda(n)$ and $\mathbf{w}_{\mathrm{a}}$ are given by

$$\lambda(n) = |\hat{X}(n)|^2, \mathbf{w}_a^{\mathrm{ML}} = (\mathbf{B}^H \mathbf{\Phi}_{\tilde{y}} \mathbf{B} + \alpha \mathbf{I}_{M-1})^{-1} \mathbf{B}^H \mathbf{\Phi}_{\tilde{y}} \mathbf{w}_{\mathrm{q}}, \tag{7}$$

where $\mathbf{\Phi}_{\tilde{y}} = \sum_n \mathbf{y}(n)\mathbf{y}^H(n)/\lambda(n)$ is the variance-weighted sample covariance matrix of $\mathbf{y}(n)$, and $\mathbf{I}_{M-1}$ is an $(M-1) \times (M-1)$ identity matrix. The parameters $\lambda(n)$ and $\mathbf{w}_{\mathrm{a}}$ can be initialized to $|Y_1(n)|^2$ and a zero vector, respectively.

After inspection of (7), we observe that the proposed ML-AIC can alleviate the speech distortion by minimizing the power of output for noise-dominant periods. Specifically, the noisy signal $\mathbf{y}(n)$ is emphasized in $\mathbf{\Phi}_{\tilde{y}}$ when $\lambda(n)$ is small, i.e., the noise dominant periods. As a consequence, $\lambda(n)$ can be considered as a soft voice activity detector (SVAD) that exploits the non-Gaussianity of the speech signal.

Note that the optimal solution of the conventional AIC is derived by minimizing the power of $\hat{X}(n)$ [3], [4], which is denoted as Minimum Power AIC (MP-AIC). In contrast to MP-AIC, the proposed ML-AIC is derived by forcing the beamformer's output as non-Gaussian as possible. A SVAD $\lambda(n)$ is intrinsically incorporated into the iterative update rules for the coefficients of the AIC $\mathbf{w}_a^{\mathrm{ML}}$, which helps to reduce contributions of speech dominant periods and thus alleviate the severe speech distortion.

B. Online Algorithm

We now develop an online version of the ML-AIC, which renders a low-latency implementation and is capable of updating the AIC coefficients in a frame-by-frame fashion. The online update rules for the ML-AIC can be derived with the stochastic gradient-based method. However, the performance of the adaptive filter may be limited by this means, e.g., the slow convergence rate and poor tracking capability [25]. To avoid these problems, we adapt the merit of the recursive least square (RLS), i.e., exponentially weighting the loss function at each time slot, to realize the online update of the ML-AIC. By doing so, the cost function $\mathcal{J}$ is thus modified as follows

$$\mathcal{J}(\mathbf{w}_{\mathrm{a}}(n), \lambda) = \sum_{l=1}^{n} \gamma^{n-l} \left(\ln(\lambda(l)) + \frac{|Y_d(l) - \mathbf{w}_{\mathrm{a}}^H(n)\mathbf{u}(l)|^2}{\lambda(l)} \right) + \alpha \mathbf{w}_{\mathrm{a}}^H(n)\mathbf{w}_{\mathrm{a}}(n), \tag{8}$$

where $0 < \gamma < 1$ is the forgetting factor. By setting the partial derivatives of $\mathcal{J}(\mathbf{w}_{\mathrm{a}}(n), \lambda)$ with respect to λ and $\mathbf{w}_{\mathrm{a}}$ to zero, the update rules read

$$\lambda(n) = |\hat{X}(n)|^2, \mathbf{w}_{\mathrm{a}}^{\mathrm{ML}}(n) = \mathbf{\Psi}^{-1}(n)\mathbf{p}(n), \tag{9}$$

where

$$\mathbf{\Psi}(n) = \gamma \mathbf{\Psi}(n-1) + 1/\lambda(n)\mathbf{u}(n)\mathbf{u}^H(n) + \alpha \mathbf{I}_{M-1}, \tag{10}$$

$$\mathbf{p}(n) = \gamma \mathbf{p}(n-1) + 1/\lambda(n)\mathbf{u}(n)Y_{\mathrm{d}}^*(n). \tag{11}$$

We find that $\mathbf{\Psi}(n)$ is the variance-weighted covariance matrix with diagonal loading, and $\mathbf{p}(n)$ is the variance-weighted cross correlation vector between $Y_d(n)$ and $\mathbf{u}(n)$. Although the BM cannot completely suppress the target speech in practice, the proposed online ML-GSC can alleviate the speech distortion due to the incorporation of the adaptive weight factor $1/\lambda(n)$. Specifically, since the term $1/\lambda(n)$ is very small when the speech is dominant, the corresponding contribution of $\mathbf{u}(n)$ and $Y_d(n)$ to the adaptation of the AIC is adaptively reduced. As a consequence, it is expected that the severe speech distortion can be mitigated.

For real-time applications, it is prohibitive to directly compute the matrix inversion $\mathbf{\Psi}^{-1}(n)$ in (9). However, $\mathbf{\Psi}^{-1}(n)$ can be effectively calculated using Sherman-Morrison inverse formula as

$$\begin{aligned}\mathbf{\Psi}^{-1}(n) &\approx \left(\gamma \mathbf{\Psi}(n-1) + 1/\lambda(n)\mathbf{u}(n)\mathbf{u}^H(n)\right)^{-1} \\ &= \frac{1}{\gamma}\left(\mathbf{\Psi}^{-1}(n-1) - \frac{\mathbf{u}_\Psi(n)\mathbf{u}_\Psi^H(n)}{\gamma\lambda(n) + \mathbf{u}^H(n)\mathbf{u}_\Psi(n)}\right),\end{aligned} \tag{12}$$

where $\mathbf{u}_\Psi(n) = \mathbf{\Psi}^{-1}(n-1)\mathbf{u}(n)$. Substituting (10), (11), and (12) into (9), we obtain the iterative rules of $\mathbf{w}_a^{\mathrm{ML}}(n)$

$$\mathbf{w}_a^{\mathrm{ML}}(n) = \mathbf{w}_{\mathrm{a}}^{\mathrm{ML}}(n-1) + \mathbf{k}(n)\xi^*(n), \tag{13}$$

$$\xi^*(n) = Y_{\mathrm{d}}^*(n) - \mathbf{u}^H(n)\mathbf{w}_a^{\mathrm{ML}}(n-1), \tag{14}$$

$$\mathbf{k}(n) = \frac{\mathbf{u}_\Psi(n)}{\gamma'(n) + \mathbf{u}^H(n)\mathbf{u}_\Psi(n)}, \tag{15}$$

and $\gamma^{'}(n) = \gamma\lambda(n)$. Note that the only difference between the RLS-based ML-AIC and the RLS-based MP-AIC is that the forgetting factor of the former is automatically adjusted. With the careful inspection of (13) and (15), we find that $\lambda(n)$ actually plays an important role in adjusting the forgetting factor, and thus adaptively control the adaptation of $\mathbf{w}_{\mathrm{a}}$. Specifically, if that target speech variance $\lambda(n)$ is large, i.e., the speech is dominant at time n, the actual forgetting factor $\gamma^{'}(n)$ is increased and the adaptation of $\mathbf{w}_{\mathrm{a}}$ is slowed down, which helps to mitigate the performance degradation caused by the speech leakage from the BM.

The accurate estimation of the target speech variance $\lambda(n)$ is crucial for the proposed ML-AIC. Supposing that the AIC has converged to some degree, the target speech variance $\lambda(n)$ can be well approximated by the variance of the output of the GSC. Thus, the variance $\lambda(n)$ can be estimated recursively as

$$\hat{\lambda}(n) = \delta\hat{\lambda}(n-1) + (1-\delta)|\mathbf{w}_{\mathrm{GSC}}^H(n)\mathbf{y}(n)|^2, \tag{16}$$

where δ is a smoothing factor with $\delta \in [0, 1]$. To make the estimation more robust, we initialize the estimation of the variance $\lambda(n)$ at time n as follows

$$\tilde{\lambda}(n) = \max\left(\delta\hat{\lambda}(n-1) + (1-\delta)|\mathbf{w}_{\mathrm{GSC}}^H(n-1)\mathbf{y}(n)|^2, \epsilon\right), \tag{17}$$

where ϵ is a small positive constant. As shown in (13) to (17), the proposed online ML-AIC algorithm can be operated in a frame-by-frame processing way and the computational complexity at each frequency bin is $\mathcal{O}(M^2)$. The computational burden can be further reduced by resorting to the low-complexity implementation of the RLS [26], [27], but it is out of the scope of this paper.

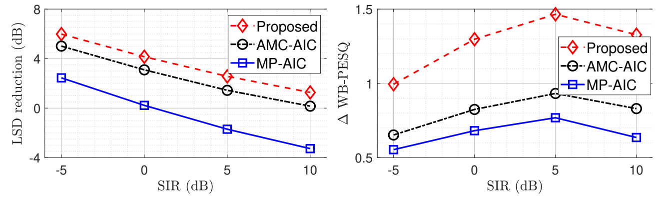

Fig. 1. Performance comparison under different SIRs. (a) LSD reduction, (b) ΔWB-PESQ. Conditions: T_{60} = 360 ms.

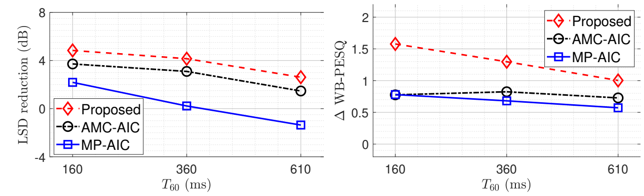

Fig. 2. Performance comparison under different T_{60}s. (a) LSD reduction, (b) ΔWB-PESQ. Conditions: SIR = 0 dB.

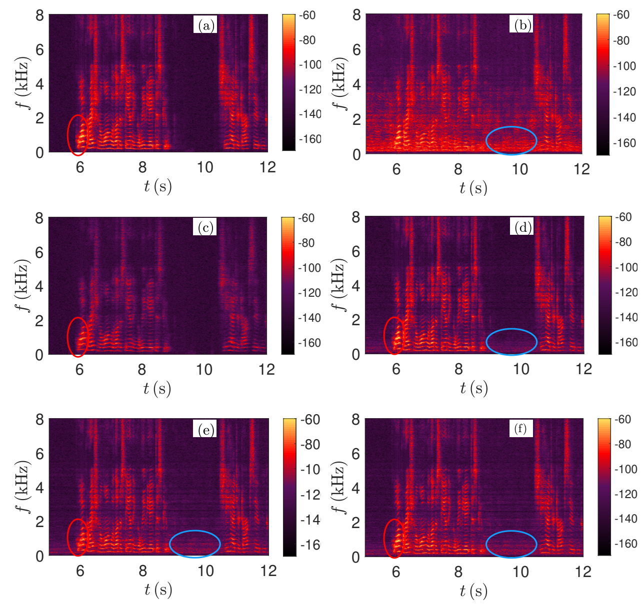

Fig. 3. Spectrograms of (a) clean speech at the reference channel, (b) noisy signal at the reference channel, (c) speech leakage at noise reference, and signals enhanced by the (d) proposed ML-AIC, (e) MP-AIC, and (f) AMC-AIC. Conditions: T_{60} = 360 ms and SIR = 5 dB.

IV. Simulations

We now conduct computer simulation to verify the performance of the proposed online ML-AIC. Specifically, we compare the MP-AIC without adaptation control [6], the AMC-AIC based on the hard-decision scheme [2] and the proposed ML-AIC under different reverberation times T_{60} and SIRs. We consider a uniform linear array of $M = 8$ microphones with the interelement distance of $d = 0.08$ m. The target source is at 0 ° and interference is at 60 °. Both the target source and the interference are positioned 2 m from the center of the linear array. The randomly selected utterances from the TIMIT [28] are concatenated and a 5 s silence period is inserted at the beginning to form a single speech signal having 20 s duration, and there are 10 speech signals in all. The interference is the babble noise taken from the NOISEX-92 database [29]. The anechoic speech and noise signals are then convolved with the corresponding measured room impulse responses [30] to generate microphone signals. The spatially uncorrelated Gaussian noise is also added with an SNR (signal-to-noise ratio) of 30 dB.

The sampling rate is 16 kHz, and we use a 64 ms hamming window with 75 % overlap. The RTFs are estimated by the covariance whiten method [34] and then utilized to construct the FB and the BM. All AICs are updated based on the RLS, and $\gamma = 0.99$. The threshold of the AMC-AIC is empirically set to 0.8 for all frequency bins, and the AIC is continuously updated without utilizing the AMC during the first 4 s. For the proposed ML-AIC, we use $\delta = 0.8, \alpha = 10^{-4}, \mathbf{\Psi}^{-1}(0) = \mathbf{I}_{M-1}, \mathbf{p}(0) = [1, \cdots, 1]^T$ and $\epsilon = 10^{-8}$.

The speech quality and speech distortion are measured in terms of wide-band perceptual evaluation of speech quality (WB-PESQ) [31] and log-spectral distance (LSD) [32]. The measures are evaluated by comparing the clean speech signal received at the first microphone, namely $X_1(n)$ with its estimate $\hat{X}(n)$. Compared to the LSD, the WB-PESQ complies better with the subjective speech quality [33]. All measures are computed by averaging the results regarding the 10 speech signals as references.

A. Under Different SIRs

We first evaluate the performance of the GSC with different AICs under various SIRs.We use the reverberation time T_{60} = 360 ms. The reduction of the LSD and the improvement of the WB-PESQ of three algorithms are presented in Fig. 1 as a function of SIR. It is observed that under different SIRs, the proposed ML-AIC with the inherent incorporation of the adaptive forgetting factor can greatly alleviate the speech distortion problem encountered in the conventional GSC and consistently outperforms the MP-AIC and the AMC-AIC.

B. Under Different Reverberation Times

We then compare the performance of GSC with different AICs under different T_{60}s, where we use SIR = 0 dB. Fig. 2 depicts the performance of the three algorithms measured by the reduction of the LSD and the improvement of the WB-PESQ. It is apparent that the proposed ML-AIC performs better than the other two AICs under different T_{60}s. As expected, the performance of GSC with different AICs is degraded and the performance difference is less significant as T_{60} increases. The reason may be that the MTF approximation adopted in (1) is not accurate for a highly reverberant environment, and the CTF-GSC may be adopted to address this problem as mentioned before.

C. Spectrogram Examples

We show the spectrograms of the various signals in Fig. 3 for T_{60} = 360 ms and SIR = 5 dB. According to Fig. 3(d)–Fig. 3(f), it is apparent that all AICs can reduce the babble noise. But the proposed method suppresses more noise and obtains less speech distortion as compared to the other two methods. Specifically, during periods around 6 s, the MP-AIC encounters speech distortion due to a large amount of speech leakage as shown clearly in Fig. 3(c) and uncontrolled adaptations of the AIC. In contrast, the proposed ML-AIC and AMC-AIC greatly alleviate speech distortion. Furthermore, during noise-only periods from 9 to 10 s, the proposed ML-AIC method provides more aggressive noise reduction compared to the other two AICs.

V. Conclusion

We have proposed a new update rule for the AIC in the GSC beamformer applied for speech enhancement. The optimal AIC has been derived as a maximum likelihood solution by assuming that the output of the GSC beamformer follows a complex Gaussian distribution with time-varying variances. We show that the update of the proposed ML-AIC is automatically controlled by the GSC output's variance, and the online version of the ML-AIC can be regarded as the RLS-based AIC with a variable forgetting factor. Experimental results substantiate the effectiveness of the proposed method.

References

[1] L. Griffiths and C. Jim, "An alternative approach to linearly constrained adaptive beamforming," *IEEE Trans. Antennas and Propagation*, vol. 30, no. 1, pp. 27–34, Jan. 1982.

[2] O. Hoshuyama and A. Sugiyama, "Robust adaptive beamforming," in *Microphone Arrays*, M. Brandstein and D. Ward, Eds. Springer, 2001, ch. 5, pp. 87–106.

[3] A. Spriet, M. Moonen, and J. Wouters, "Spatially pre-processed speech distortion weighted multi-channel Wiener filtering for noise reduction," *Signal Process.*, vol. 84, no. 12, pp. 2367–2387, 2004.

[4] A. Spriet, M. Moonen, and J. Wouters, "Robustness analysis of multichannel wiener filtering and generalized sidelobe cancellation for multimicrophone noise reduction in hearing aid applications," *IEEE Trans. Speech, Audio Process.*, vol. 13, no. 4, pp. 487–503, Jul. 2005.

[5] 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.

[6] S. Gannot, D. Burshtein, and E. Weinstein, "Signal enhancement using beamforming and nonstationarity with applications to speech," *IEEE Trans. Signal Process.*, vol. 49, no. 8, pp. 1614–1626, Aug. 2001.

[7] B.-J. Yoon, I. Tashev, and A. Acero, "Robust adaptive beamforming algorithm using instantaneous direction of arrival with enhanced noise suppression capability," in *Proc. IEEE ICASSP*, vol. 1, Honolulu, HI, USA, Jun. 2007, pp. I–133–I–136.

[8] R. Talmon, I. Cohen, and S. Gannot, "Convolutive transfer function generalized sidelobe canceler," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 7, pp. 1420–1434, Sep. 2009.

[9] S. Markovich-Golan, S. Gannot, and I. Cohen, "A sparse blocking matrix for multiple constraints GSC beamformer," in *Proc. IEEE ICASSP*, Kyoto, Japan, Aug. 2012, pp. 197–200.

[10] Y. Zheng, K. Reindl, and W. Kellermann, "Analysis of dual-channel ICA-based blocking matrix for improved noise estimation," *EURASIP J. Adv. Signal Process.*, vol. 2014, no. 1, pp. 1–24, 2014.

[11] L. Wang, T. Gerkmann, and S. Doclo, "Noise power spectral density estimation using MaxNSR blocking matrix," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 23, no. 9, pp. 1493–1508, Sep. 2015.

[12] O. Hoshuyama and A. Sugiyama, "An adaptive microphone array with good sound quality using auxiliary fixed beamformers and its DSP implementation," in *Proc. IEEE ICASSP*, vol. 2, Phoenix, AZ, USA, Aug. 1999, pp. 949–952.

[13] W. Herbordt and W. Kellermann, "Computationally efficient frequency-domain robust generalized sidelobe canceller," in *Proc. IWAENC*, 2001, pp. 51–55.

[14] M.-S. Choi, C.-H. Baik, Y.-C. Park, and H.-G. Kang, "A soft-decision adaptation mode controller for an efficient frequency-domain generalized sidelobe canceller," in *Proc. IEEE ICASSP*, vol. 4, Honolulu, HI, USA, Jun. 2007, pp. IV–893–IV–896.

[15] W. Herbordt, H. Buchner, S. Nakamura, and W. Kellermann, "Multichannel bin-wise robust frequency-domain adaptive filtering and its application to adaptive beamforming," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 4, pp. 1340–1351, May. 2007.

[16] K. Kumatani, J. McDonough, B. Rauch, P. N. Garner, W. Li, and J. Dines, "Maximum kurtosis beamforming with the generalized sidelobe canceller," in *Proc. Interspeech*, Brisbane, Australia, Sep. 2008, pp. 423–426.

[17] K. Kumatani, J. McDonough, B. Rauch, D. Klakow, P. N. Garner, and W. Li, "Beamforming with a maximum negentropy criterion," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 17, no. 5, pp. 994–1008, Jul. 2009.

[18] K. Kumatani, T. Gehrig, U. Mayer, E. Stoimenov, J. McDonough, and M. Wölfel, "Adaptive beamforming with a minimum mutual information criterion," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 8, pp. 2527–2541, Nov. 2007.

[19] K. Reindl, S. Meier, H. Barfuss, and W. Kellermann, "Minimum mutual information-based linearly constrained broadband signal extraction," *IEEE/ACM Trans. Audio, Speech, Lang. Process.*, vol. 22, no. 6, pp. 1096–1108, Jun. 2014.

[20] K. Kumatani, J. McDonough, and B. Raj, "Block-wise incremental adaptation algorithm for maximum kurtosis beamforming," in *Proc. IEEE WASPAA*, New Paltz, NY, USA, Oct. 2011, pp. 229–232.

[21] C. Févotte and J.-F. Cardoso, "Maximum likelihood approach for blind audio source separation using time-frequency Gaussian source models," in *Proc. IEEE WASPAA*, New Paltz, NY, USA, Oct. 2005, pp. 78–81.

[22] T. Nakatani, B.-H. Juang, T. Yoshioka, K. Kinoshita, M. Delcroix, and M. Miyoshi, "Speech dereverberation based on maximum-likelihood estimation with time-varying Gaussian source model," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 16, no. 8, pp. 1512–1527, 2008.

[23] B. J. Cho, J.-M. Lee, and H.-M. Park, "A beamforming algorithm based on maximum likelihood of a complex Gaussian distribution with time-varying variances for robust speech recognition," *IEEE Signal Process. Lett.*, vol. 26, no. 9, pp. 1398–1402, Sep. 2019.

[24] J. S. Erkelens, R. C. Hendriks, R. Heusdens, and J. Jensen, "Minimum mean-square error estimation of discrete Fourier coefficients with generalized Gamma priors," *IEEE Trans. Audio, Speech, Lang. Process.*, vol. 15, no. 6, pp. 1741–1752, 2007.

[25] S. Haykin, *Adaptive filter theory*, 5th ed. New Jersey, USA: Pearson Education, Inc., 2014.

[26] Y. V. Zakharov, G. P. White, and J. Liu, "Low-complexity RLS algorithms using dichotomous coordinate descent iterations," *IEEE Trans. Signal Process.*, vol. 56, no. 7, pp. 3150–3161, Jul. 2008.

[27] G.-O. Glentis, K. Berberidis, and S. Theodoridis, "Efficient least squares adaptive algorithms for FIR transversal filtering," *IEEE Signal Process. Magazine*, vol. 16, no. 4, pp. 13–41, 1999.

[28] V. Zue, S. Seneff, and J. Glass, "Speech database development at MIT: TIMIT and beyond," *Speech communication*, vol. 9, no. 4, pp. 351–356, 1990.

[29] A. Varga and H. J. Steeneken, "Assessment for automatic speech recognition: II. NOISEX-92: A database and an experiment to study the effect of additive noise on speech recognition systems," *Speech communication*, vol. 12, no. 3, pp. 247–251, 1993.

[30] E. Hadad, F. Heese, P. Vary, and S. Gannot, "Multichannel audio database in various acoustic environments," in *Proc. IWAENC*, Juan-les-Pins, France, Sep. 2014, pp. 313–317.

[31] ITU-T Recommendation P. 862. 2, *Wideband extension to Recommendation P.862 for the assessment of wideband telephone networks and speech codecs*, 2007.

[32] I. Cohen and S. Gannot, "Spectral enhancement methods," *Springer Handbook of Speech Processing*, pp. 873–902, 2008.

[33] H. Gamper, C. K. Reddy, R. Cutler, I. J. Tashev, and J. Gehrke, "Intrusive and non-intrusive perceptual speech quality assessment using a convolutional neural network," in *Proc. IEEE WASPAA*, New Paltz, NY, USA, Oct. 2019, pp. 85–89.

[34] S. Markovich-Golan, S. Gannot, and W. Kellermann, "Performance analysis of the covariance-whitening and the covariance-subtraction methods for estimating the relative transfer function," in *Proc. EUSIPCO*, Rome, Italy, Sep. 2018, pp. 2499–2503.