A Robust GSC Beamforming Method for Speech Enhancement using Linear Microphone Array

Feng Ni, Yi Zhou, and Hongqing Liu
School of Communication and Information Engineering
Chongqing University of Posts and Telecommunications
Chongqing, China
chris.ni@foxmail.com; zhouy@cqupt.edu.cn; hongqingliu@cqupt.edu.cn

***Abstract*—The speech enhancement problem is studied using an improved robust generalized sidelobe canceler (GSC) beamforming algorithm based on microphone array, in the cases of speaker noise and the music interferences. The conventional GSC algorithm based on variable step size and a priori signal-to-noise ratio (SNR) algorithm is not robust under the nonstationary noise because the solution of the signal-to-noise ratio (SNR) is not given. To enhance the robustness, in this paper, a improved GSC algorithm is developed, where adaptive filter coefficients are updated based on signal output power ratio (SPR). The numerical studies including speaker noise and music noise demonstrate that the improved algorithm outperforms the traditional GSC and the GSC based on variable step size technique.**

***Index Terms*—Generalized sidelobe canceler, NLMS, SPR**

I. INTRODUCTION

In speech communication system applications such as mobile phones, hand-free communications, hearing aids, speech recognition, audio processing in virtual reality, stereo-sound systems, speech enhancement plays an important role. For past few decades, many speech enhancement algorithms have been developed to improve signal quality and listener understandability in noisy and reverberant environments. The quality of speech signal indicates clarity of the speech signal, whereas intelligibility indicates the understandability of words by listeners. Generally speaking, speech enhancement algorithms are selected based on the type of interference signal and different noisy and reverberant cases [1].

For single microphone case, speech enhancement algorithms such as spectral subtraction [2], Weiner filter algorithms [3], to name a few, are popular choices because their simplicity and guaranteed performance. They require less cost and less complexity, but these algorithms are unable to reduce the musical noise and sometimes fail to remove background noise. Recently, microphone array becomes increasingly important because it retrieves the desired clean speech signal from reverberation and noisy conditions using spatial information. In microphone array, beamforming technique is one of the multi-channel algorithms that is frequently utilized for speech enhancement.

In beamfroming methods, the generalized sidelobe canceler (GSC) is effective for the speech enhancement in a reverberating environment. In [4], Frost developed a constrained minimum power adaptive beamforming, where pure delay relates each pair of source and sensor. Griffiths and Jim [5] reconsidered Frost algorithm and introduced the generalized sidelobe canceler (GSC) solution. In that, GSC algorithm is comprised of three building blocks. The first is a Fixed beamformer Filter (FBF), which satisfies the desired constraint. The second is a blocking matrix (BM), which produces noise-only reference signal by blocking the desired signal. The third is an adaptive noise canceller (ANC) filter that attempts to cancel the noise from the fixed beamformer output. It is shown that Frost algorithm is a special case of the GSC. In [6], Hoshuyama used a three-block structure similar to the GSC, but, the blocking matrix has been modified to operate adaptively to limit the leakage of the desired signal. The studies show that the methods of Griffiths and Jim [14] lack the ability to handle the nonstationary acoustic signal.

To improve the robustness, in this paper, a new improved GSC beamforming method using microphone array for the speech enhancement is developed. The proposed method presents the ability to handle the nonstationary speech and promote the intelligibility of the speech.

The rest of the paper is organized as follows. The traditional GSC speech enhancement method is introduced in section II. The proposed approach is presented in section III. Experimental settings and comparison results on the CHIME-3 database are described in section IV. Finally, section V concludes this paper.

II. SIGNAL MODEL AND GSC MODEL

A. Signal Model of Linear Microphone Array

In this work, an array of M sensors in a noisy and reverberant environment is considered, depicted in Fig.1. The received signals generally include two components. The first is a desired speech signal, and the second is the interference including stationary and nonstationary (transient) components. The goal of speech enhancement is to reconstruct the speech component from the received signals and suppress the interference. Let $s(t)$ denote the desired source signal, $a_{\mathrm{m}}(t)$ represent the room impulse response (RIR) of the m-th sensor to the desired source, and $n_{\mathrm{m}}(t)$ denote the noise component at the m-th sensor, and the observed signal at the m-th sensor $(m = 1, \cdots, M)$ is given by

$$z_{\mathrm{m}}(t) = a_{\mathrm{m}}(t) * s(t) + n_{\mathrm{m}}(t), \tag{1}$$

978-1-7281-1817-8/19/$31.00 ©2019 IEEE

where $*$ denotes convolution, and $n_{\mathrm{m}}(t)$ contains both directional noise component and diffused noise component.

The observed signals are now divided into overlapping frames by a use of Hanning window function and after short-time Fourier transform (STFT), Equation (1) in the time-frequency domain is

$$\mathbf{Z}(k,\ell) = \mathbf{A}(k)S(k,\ell) + \mathbf{N}(k,\ell), \tag{2}$$

where

$$\mathbf{Z}(k,\ell) = [Z_1(k,\ell), Z_2(k,\ell)\cdots Z_M(k,\ell)], \tag{3}$$

$$\mathbf{A}(k,\ell) = [A_1(k,\ell), A_2(k,\ell)\cdots A_M(k,\ell)], \tag{4}$$

$$\mathbf{N}(k,\ell) = [N_1(k,\ell), N_2(k,\ell)\cdots N_M(k,\ell)], \tag{5}$$

where ℓ is the frame index and k represents the frequency bin index.

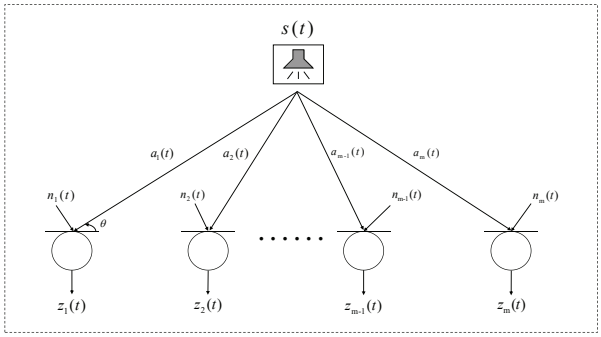

Fig. 1: Signal model for linear microphone array

B. Frequency-Domain Generalized Sidelobe Canceller (GSC)

The GSC beamformer consists of three parts and the first is the fixed beamformer. The simplest and yet the most widely used the fixed beamformer is the delay-and-sum (DSB), and in the DSB, using the first microphone as the reference microphone, the steering vector $\mathbf{A}(k)$ is

$$\mathbf{A}(k) = [1, e^{-j\frac{2\pi}{k}\tau_2}, e^{-j\frac{2\pi}{k}\tau_3}\cdots e^{-j\frac{2\pi}{k}\tau_{\mathrm{m}}}], \tag{6}$$

where $\tau_2, \cdots \tau_{\mathrm{m}}$ are the relative delays between each microphone and the reference microphone. With the steering vector, the FBF is

$$\mathbf{W}_0(k) = (1, e^{j\frac{2\pi}{k}\tau_2}, e^{j\frac{2\pi}{k}\tau_3}\cdots e^{j\frac{2\pi}{k}\tau_{\mathrm{m}}}). \tag{7}$$

The blocking matrix $\mathbf{\Gamma}(k)$ that blocks the desired signal is [14]

$$\mathbf{\Gamma}(k) = \begin{pmatrix} -e^{-j\frac{2\pi}{k}\tau_2} & -e^{-j\frac{2\pi}{k}\tau_3} & \cdots & -e^{-j\frac{2\pi}{k}\tau_m} \\ 1 & 0 & \cdots & 0 \\ 0 & 1 & \cdots & 0 \\ & \cdots & \ddots & \\ 0 & 0 & \cdots & 1 \end{pmatrix} \tag{8}$$

Utilizing the fixed beamformer filter, the output of $\mathbf{Z}(k,\ell)$, denoted by $Y_{\mathrm{FBF}}(k,\ell)$, is

$$Y_{\mathrm{FBF}}(k,\ell) = \mathbf{W_0}^{\mathrm{H}}(k)\mathbf{Z}(k,\ell). \tag{9}$$

The output of BM is called noise reference signal $U(k,\ell)$, given by

$$U(k,\ell) = \mathbf{\Gamma}^{\mathrm{H}}(k)\mathbf{Z}(k,\ell). \tag{10}$$

The final part is a multichannel ANC, denoted by $\mathbf{G}(k,l)$ and it is defined by

$$\mathbf{G}(k,l) = (G_2(k,\ell), G_3(k,\ell)\cdots G_M(k,\ell)). \tag{11}$$

Using (10) and (11), the output of ANC $Y_{\mathrm{ANC}}(k,\ell)$ is

$$Y_{\mathrm{ANC}}(k,\ell) = \mathbf{G}^{\mathrm{H}}(k,l)\mathbf{\Gamma}^{\mathrm{H}}(k)\mathbf{Z}(k,\ell). \tag{12}$$

Subtracting $Y_{\mathrm{ANC}}(k,\ell)$ from the output of the FBF yields

$$Y(k,\ell) = Y_{\mathrm{FBF}}(k,\ell) - Y_{\mathrm{ANC}}(k,\ell), \tag{13}$$

Substituting (9) and (12) into (13) produces the final output, given by

$$Y(k,\ell) = [\mathbf{W}_0{}^{\mathrm{H}}(k) - \mathbf{G}(k,l)\mathbf{\Gamma}^{\mathrm{H}}(k)]\mathbf{Z}(k,\ell), \tag{14}$$

where $Y(k,\ell)$ is the output of the constrained beamformer.

It is worth mentioning that, even when the best BM is utilized, $U(k,\ell)$ is noisy, and the desired speech components may leak into the noise reference signals. The residual noise in $Y_{\mathrm{FBF}}(k,\ell)$ can then be reduced by properly adjusting the filter $\mathbf{G}(k,l)$ by the minimum output power criterion, which is the classical multichannel noise cancellation problem. In practice, this optimization problem is often solved by using the normalized LMS (NLMS) [15] algorithm. The filter $\mathbf{G}(k,\ell+1)$ is updated as

$$\mathbf{G}(k,\ell+1) = \begin{cases} \mathbf{G}(k,\ell) + \mu\frac{U(k,\ell)Y'(k,\ell)}{P_{\mathrm{est}}(k,\ell)} & H_{0s}\mathrm{true}, \\ \mathbf{G}(k,\ell), & \mathrm{otherwise}, \end{cases} \tag{15}$$

$$P_{\mathrm{est}}(k,\ell) = \alpha_{\mathrm{p}}P_{\mathrm{est}}(k,\ell-1) + (1-\alpha_{\mathrm{p}}) \parallel U(k,\ell)\parallel^2, \tag{16}$$

where $P_{\mathrm{est}}(k,\ell)$ represents the power of the noise reference signals, μ is a step size controlling the convergence rate and α_{p} is forgetting factor, respectively. The $Y'(k,\ell)$ represents conjugation operation of $Y(k,\ell)$, and H_{0s} indicates the absence of transients.

III. PROPOSED METHOD

A. Classical Algorithm of ANC

Since the desired signal will leak from the BM output, this leads to the distortion of the desired signal. To solve this problem, lots of approaches have been developed. Greenberg [9] proposed a cross-correlation approach, but it inevitably introduces music noise. In [7], Hoshuyama constrained adaptive blocking matrix (ABM) and leakage adaptive filter (LAF) to reduce cancellation of desired signals. However, it is shown that ABM still suppresses desired signal while noise is noticeable. Therefore, the overall performance is affected. In [7], the signal-to-noise ratio (SNR) is defined by

$$s_{\mathrm{SNR}}(k) = \frac{p_{\mathrm{F}}(k)}{p_{\mathrm{B}}(k)}, \tag{17}$$

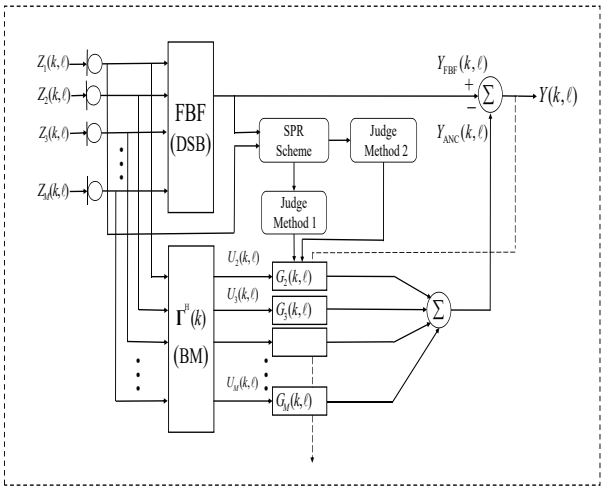

Fig. 2: The framework of the proposed improved robust GSC for speech enhancement

where $p_{\mathrm{F}}(k)$ and $p_{\mathrm{B}}(k)$ are FBF output power and BM output power, respectively. Using the output powers of FBF and ABM, SNR is utilized to control the updates of ABM and ANC. When the SNR is greater than a threshold, the filter coefficient of ABM is updated, and when the SNR is less than a threshold, the ANC of filter coefficient is updated. In [11], Scalart proposed to control filter update using variable step factor and a priori signal-to-noise ratio (SNR_{prior}), given by

$$\mu_{\varepsilon}=\frac{1}{SNR_{\mathrm{prior}}(k)+1}. \tag{18}$$

However, the specific method of SNR_{priori} is not provided. In [10], Cohen proposed a signal detection method based on instantaneous beam reference ratio, which improved the accuracy of signal decision in the nonstationary noise environment. Although the above methods are effective, they are usually adopted in the post-filtering algorithm and they are rarely applied in the ANC part. It is well known that, the traditional ANC uses the NLMS algorithm, and the convergence rate and robustness of the adaptive algorithm used by ANC have a significant impact on the noise reduction effect of GSC.

B. Robust GSC-SPR Beamforming Method

From the above description and analysis, a robust GSC beamforming method is developed, and it is based on signal output power ratio (SPR) to adaptively update the filter coefficients. The overall algorithm structure diagram is depicted in Fig. 2.

Rewritting the FBF output power $p_{\mathrm{F}}(k)$ as $P_{\mathrm{f\_o}}(k,\ell)$, and the power of FBF is calculated based on the average of frenquncy indexes from k=32 to k=256, given by,

$$P_{\mathrm{f\_o_1}}(k,\ell)=\alpha_s P_{\mathrm{f\_o_1}}(k,\ell)+(1-\alpha_s)\sum_{k=32}^{256}\parallel P_{\mathrm{f\_o}}(k,\ell)\|^2, \tag{19}$$

where α_s is a smoothing parameter that is chosen empirically and the $P_{\mathrm{f\_o}}(k,\ell)$ is output signal of FBF beamformer. Using smoothing operations reduce false judgement for weak speech so that speech distortion of the desired signal is alleviated. The FBF output power in all frames now calculated as

$$P_{\mathrm{f\_o}}(k,\ell)=\sum_{i=1}^{\ell}P_{\mathrm{f\_o_1}}(k,i). \tag{20}$$

With the same principle, the output power of smoothing within the constrained frequency in the desired signal is

$$P_{\mathrm{s\_f_1}}(k,\ell)=\alpha_s P_{\mathrm{s\_f_1}}(k,\ell)+(1-\alpha_s)\sum_{k=32}^{256}\parallel S_1(k,\ell)\|^2, \tag{21}$$

The first microphone output power in all frames now calculated as

$$P_{\mathrm{s\_f}}(k,\ell)=\sum_{i=1}^{\ell}P_{\mathrm{s\_f_1}}(k,i), \tag{22}$$

where $S_1(k,\ell)$ is the desired signal of the first microphone. With (20) and (22), the SPR, in this work, is defined by

$$SPR=\frac{P_{\mathrm{f\_o}}(k,\ell)}{\max(P_{\mathrm{s\_f}}(k,\ell),eps)}. \tag{23}$$

Utilizing SPR, a soft judgement is applied to derive a new step factor. That is,

$$\mu_{\mathrm{adjust}}=\begin{cases}0.01\mu, & \mathrm{if}SPR>\alpha_{\mathrm{thr}}\\ \mu & \mathrm{otherwise},\end{cases} \tag{24}$$

where α_{thr} is an empirically determined threshold value. To even better preserve the desired signal, we develop a second soft judgement, shown in Fig. 2, as follows

$$\delta_0=\begin{cases}1, & \mathrm{if}SPR>\upsilon_{\mathrm{high}}\\ 0, & \mathrm{if}SPR<\upsilon_{\mathrm{low}}\\ \frac{(\upsilon_{\mathrm{high}}-SPR)}{(\upsilon_{\mathrm{high}}-\upsilon_{\mathrm{low}})} & \mathrm{otherwise},\end{cases} \tag{25}$$

where υ_{high} and υ_{low} are empirically values. The purpose of the first soft judgement is to update the filter coefficients when noise is present, and hold (slowly) the update of the filter coefficients when there is a desired signal. The objective of the second soft judgement is to protect the desired signal of the intermediate frequency. Substituting δ_0 and μ_{adjust} to (15), a new improved NLMS algorithm is

$$\mathbf{G}(k,\ell+1)=\mathbf{G}(k,\ell)+\delta_0\mu_{\mathrm{adjust}}\frac{U(k,\ell)Y'(k,\ell)}{P_{\mathrm{est}}(k,\ell)}. \tag{26}$$

In this paper, the proposed method is named GSC-SPR and in what follows, its performances compared with other approaches are provided.

IV. SIMULATION

A. Experimental setting

In this section, the performance of GSC-SPR algorithm is evaluated in the presence of interfering speaker and music noise. The comparisons are conducted with two methods, namely the traditional NLMS of GSC and the GSC based on variable step size and a priori signal-to-noise ratio (SNR) algorithm [11], termed as GSC-Steps. Utterances of desired

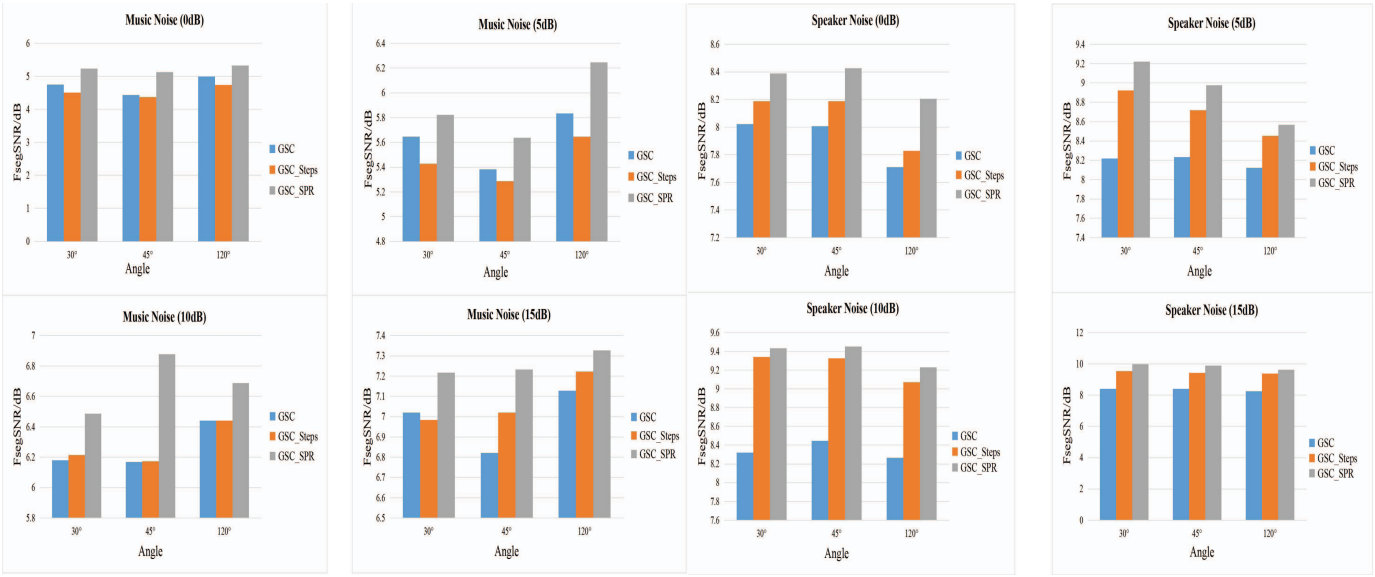

Fig. 4: The output FsegSNR (dB) obtained by three algorithms.

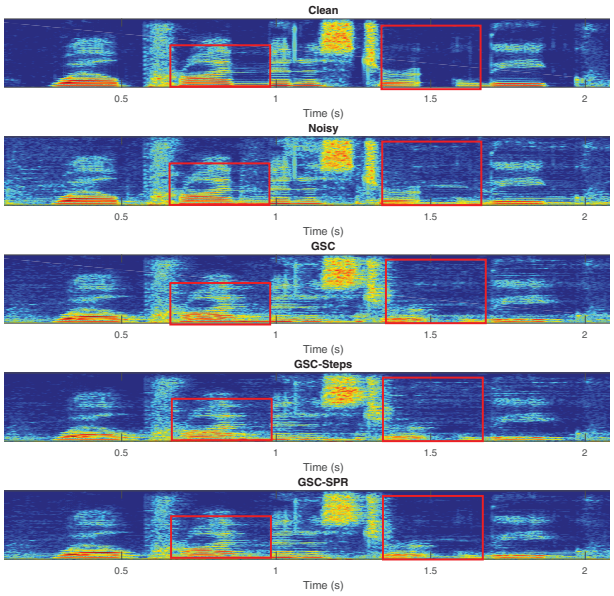

Fig. 3: Spectrograms of an utterance corrupted by 10dB music noise and its enhanced results.

and interfering speakers are taken from the WSJ0 corpus of CHiME-3 data set. The sampling rate for all speech signals and noises are 16 KHz and the frame size is set to 512, Hamming window and 50% overlap between adjacent frames are applied. For GSC-SPR, some empirically values are chosen as $\mu = 0.05$, $\alpha_{\mathrm{p}} = 0.95$, $\alpha_s = 0.9$, $\alpha_{\mathrm{thr}} = 0.38$, $\upsilon_{\mathrm{high}} = 0.5$, $\upsilon_{\mathrm{low}} = 0.2$. In our experiments, a uniform linear microphone array with four microphones is used, where the inter-microphone distance is 3.5cm. The reverberation is generated in a rectangular room and the reverberation time is

TABLE I: PESQs obtained by different approaches.

Noise (0dB)	**Speaker noise**	**Music noise**
Angle (30°, 45°,120°)	**PESQ**	**PESQ**
GSC	1.340 1.347 1.334	1.330 1.317 1.300
GSC-Steps	1.362 1.369 1.351	1.335 1.326 1.310
GSC-SPR	**1.488 1.492 1.487**	**1.458 1.452 1.438**

Noise (5dB)	**Speaker noise**	**Music noise**
Angle (30°, 45°,120°)	**PESQ**	**PESQ**
GSC	1.356 1.355 1.334	1.339 1.331 1.336
GSC-Steps	1.385 1.382 1.366	1.347 1.348 1.337
GSC-SPR	**1.558 1.546 1.532**	**1.551 1.548 1.536**

Noise (10dB)	**Speaker noise**	**Music noise**
Angle (30°, 45°,120°)	**PESQ**	**PESQ**
GSC	1.361 1.337 1.357	1.354 1.349 1.353
GSC-Steps	1.393 1.391 1.373	1.361 1.350 1.353
GSC-SPR	**1.566 1.558 1.541**	**1.545 1.539 1.540**

Noise (15dB)	**Speaker noise**	**Music noise**
Angle (30°, 45°,120°)	**PESQ**	**PESQ**
GSC	1.363 1.358 1.340	1.347 1.344 1.337
GSC-Steps	1.390 1.386 1.368	1.364 1.364 1.369
GSC-SPR	**1.616 1.594 1.572**	**1.576 1.572 1.582**

0.3s, 4m wide by 5m long by 3m high. The linear array was positioned on a table at the center of the room. Additionally, the desired source is at 90° of the microphone array, and one interference source are located at different directions of 30°, 45°, 120°. The input SNRs are 0dB, 5dB, 10dB and 15dB, respectively. The performance of different algorithms is evaluated in terms of speech frequency segment SNR (FsegSNR) [12] and perceptual evaluation of speech quality (PESQ) [13].

B. Experimental results

The Fig. 4 shows the frequency segmental SNR results of the enhanced speech. It is seen that the direction of the

interference signal affects the performance of three algorithms, and GSC-SPR outperforms others in terms of FsegSNRs under different noise environments. It is also noticed that in the case of low SNR, the performance is less satisfactory and still needs to be improved.

At different input SNRs, PESQs of different approaches are summarized in TABLE I based on the average of 30 different data sets. In the cases of music noise and speaker noise, the PESQ increases as input SNRs increase and compared with the basic GSC algorithm, the improvement is obvious. In addition to that, when input SNR is low, say 0dB, the PESQ is slightly improved, which indicates the proposed method still has limitations in the low SNR scenario. This also agrees with the findings in the FsegSNR study. Finally, the spectrograms of one example utterances corrupted by 10dB music noise and its enhanced results obtained by different approaches are shown in Fig. 3. It can be obviously seen that the proposed method achieves a better noise reduction. For a better illustration, in Fig. 3, two areas are emphasized to show the differences of noise reduction.

V. CONCLUSION

A improved robust GSC beamforming method using microphone array is proposed in this paper. To better remove the noise, based on the designed SPR, a variable step size scheme in adaptive filter is developed. The improved filter provides the ability to preserve the desired signal and at the same time, to better suppress the noise. The experimental results demonstrate that GSC-SPR algorithm produces a better performance than traditional algorithm in both speaker noise and music noise scenarios, under the reverberation environment. In the future, an attempt to improve performance in the low SNR environment will be conducted.

REFERENCES

[1] J. Benesty, M. M. Sondhi, and Y. A. Huang. *Introduction to Speech Processing*. Springer Handbook of Speech Processing. Springer, Berlin, Heidelberg, 2008: 1-4.

[2] K. Paliwal, K. Wjcicki, and B. Schwerin. Single-channel speech enhancement using spectral subtraction in the short-time modulation domain. *Speech communication*, 2010, 52(5): 450-475.

[3] J. S. Lim, A. V. Oppenheim. Enhancement and bandwidth compression of noisy speech. *Proceedings of the IEEE*, 1979, 67(12): 1586-1604.

[4] O. L. Frost An algorithm for linearly constrained adaptive array processing. *Proceedings of the IEEE*, 1972, 60(8): 926-935.

[5] L. Griffiths, C. W. Jim An alternative approach to linearly constrained adaptive beamforming. *IEEE Transactions on Antennas and Propagation*, 1982, 30(1): 27-34.

[6] O. Hoshuyama, A. Sugiyama, A. Hirano A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Transactions on Signal Processing*, 1999, 47(10): 2677-2684.

[7] O.Hoshuyama, A. Sugiyama, A.Hirano A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters. *IEEE Transactions on Signal Processing*, 1999, 47(10): 2677-2684

[8] O.Hoshuyama, B.Begasse, A.Sugiyama, et al. A real time robust adaptive microphone array controlled by an SNR estimate, *Proceedings of the 1998 IEEE International Conference on Acoustics, Speech and Signal Processing*, ICASSP'98 (Cat. No. 98CH36181). IEEE, 1998, 6: 3605-3608.

[9] J. E. Greenberg, P. M. Zurek. Evaluation of an adaptive beamforming method for hearing aids. *The Journal of the Acoustical Society of America*, 1992, 91(3): 1662-1676.

[10] I. Cohen, B.Berdugo. Multichannel signal detection based on the transient beam-to-reference ratio. *IEEE Signal Processing Letters*, 2003, 10(9): 259-262.

[11] L. Lepauloux, P.Scalart, C.Marro. Computationally efficient and robust frequency-domain GSC. *12th IEEE International Workshop on Acoustic Echo and Noise Control*. 2010.

[12] J. M. Tribolet, P. Noll, McDermott B, et al. A study of complexity and quality of speech waveform coders. *ICASSP'78. IEEE International Conference on Acoustics, Speech, and Signal Processing*, IEEE, 1978, 3: 586-590.

[13] A. W. Rix, J. G. Beerends, M. P. Hollier, et al. Perceptual evaluation of speech quality (PESQ)-a new method for speech quality assessment of telephone networks and codecs *2001 IEEE International Conference on Acoustics, Speech, and Signal Processing*. Proceedings (Cat. No. 01CH37221). IEEE, 2001, 2: 749-752.

[14] B. Widrow, S. D. Stearns. Adaptive signal processing. 1985.

[15] K. Buckley, L. Griffiths An adaptive generalized sidelobe canceller with derivative constraints. *IEEE Transactions on Antennas and Propagation*, 1986, 34(3): 311-319.