Speech Enhancement with Microphone Array Using a Multi Beam Adaptive Noise Suppressor

Mikhail Stolbov[1,2] and Alexander Lavrentyev[1(✉)]

[1] Speech Technology Center, Krasutskogo-4, St. Petersburg 196084, Russia
{stolbov, lavrentyev}@speechpro.com
[2] ITMO University, 49 Kronverkskiy Pr., St. Petersburg 197101, Russia

Abstract. This paper presents a new speech enhancement method with microphone array for the joint suppression of coherent and diffuse noise. The proposed method is based on combined technique: target and noise steering beamforming and adaptive noise suppression. The microphone array forms two beams steered in the directions of target speaker and of noise source. The signal of reference beam is used to suppress the noise in primary channel. The proposed algorithm of Adaptive Noise Suppressor (ANS) is based on the transformation of the signal spectrum of the reference channel into the noise spectrum of the main channel using noise equalizer and algorithm of dual channel spectral subtraction. The effectiveness of the proposed technique is confirmed in varying real life coherent and diffuse noise conditions. The experimental results show that proposed method is an efficient procedure for speech quality improvement in real life noisy and reverberant conditions with SNRs down to −5 dB and reverberation time up to 0.88 s.

Keywords: Speech enhancement · Adaptive interference canceller/suppressor

1 Introduction

Microphone arrays (MA) are widely used technique for speech capturing in noisy environments in all areas of speech processing. In general, the acoustic noise is a combination of diffuse and spatially coherent noise. Coherent direct path noises are produced by point acoustic sources in free space (with unhindered wave propagation). Incoherent diffuse noises are produced by remote or spatially distributed acoustic sources under conditions of reverberation and multipath wave propagation [1, 2].

The basic algorithm of MA is fixed beamforming (FBF). However, this FBF algorithm is not very efficient because part of the environment noise comes to MA output through both main lobe and sidelobes. The problem of speech enhancement is important in a high noise level conditions (sounds of audio devices indoor, art work and traffic sounds outdoor). In these cases, the signal/noise ratio (SNR) of the output signal of MA is low.

A large number of methods for suppression of coherent and diffuse noises are proposed [1–5]. For non-stationary noise environment methods of adaptive noise

A. Ronzhin et al. (Eds.): SPECOM 2016, LNAI 9811, pp. 636–644, 2016.
DOI: 10.1007/978-3-319-43958-7_77

reduction based on extraction of reference noise signal and algorithm of it suppression in noisy target signal are used. Three groups of methods are used for extraction of reference noise signal. (1): null steering beamforming (NSB) in the direction of the noise source [5]. (2): using the reference microphone placed close to the source of interference [6]. (3): forming beams (one or various) in the direction of noise sources [7–9] using the same microphone array or by an auxiliary array [10].

The disadvantage of the first group of methods is the sensitivity to steering errors (misadjustment) of the primary channel and multipath propagation of target signal [1]. The target signal leakage in reference channel results in cancellation of target signal.

A disadvantage of the second group of methods is that the placement of a microphone close to the noise source is not always possible.

The third group of forming beams (one or various) in the direction of noise sources is more robust to steering errors. Our proposed method is based on the formation of beams steered in the directions of target speaker and of noise source.

The second element of the methods of adaptive noise reduction is a signal processing algorithm. The algorithms are divided into two main classes: adaptive noise cancelling (ANC) and adaptive noise suppression (ANS) algorithms.

ANC algorithms better save the target speech signal, but their limitation is that they only suppress the coherent part of the noise. In the case of a diffuse sound field much of noise may be incoherent in main and reference channels, which weakens the effectiveness of noise suppression. ANS algorithms distort the useful signal, but they allow suppress a coherent and diffuse noise.

The main purpose of this research is to improve the microphone array algorithm of both coherent and diffuse noise reduction for speech enhancement.

2 The Proposed Method

The proposed method is based on combined technique: target and noise steering beamforming and adaptive noise suppression. The basis of beamforming is frequency-domain FBF [3]. A general block diagram for an adaptive noise cancellation/suppression (ANC/ANS) system is presented in Fig. 1. The signals of the microphones are segmented into overlapping frames with 50 % overlap. Then Hann window is applied on each segment and a set of Fourier coefficients $X_n(\omega, k)$ using short-time fast Fourier transform (STFT) is generated.

The signals of main and reference beams are calculated as follows:

$$X(\omega,k) = \boldsymbol{D}^T(\theta_M,\omega)\boldsymbol{X}(\omega,k) \qquad R(\omega,k) = \boldsymbol{D}^T(\theta_R,\omega)\boldsymbol{X}(\omega,k), \tag{1}$$

where ω – is the frequency, k – is the frame time index, θ_M,θ_R – are the angles of the directions to the target speaker and to the noise source respectively, N – is the number of microphones, $\boldsymbol{D}(\theta,\omega) = [d_1(\theta,\omega), d_2(\theta,\omega), \ldots, d_N(\theta,\omega)]^T$ is the steering vector of the microphone array in the direction θ, $\boldsymbol{X}(\omega,k) = [X_1(\omega,k), X_2(\omega,k), \ldots, X_N(\omega,k)]^T$ is the signal vector received by the microphone array.

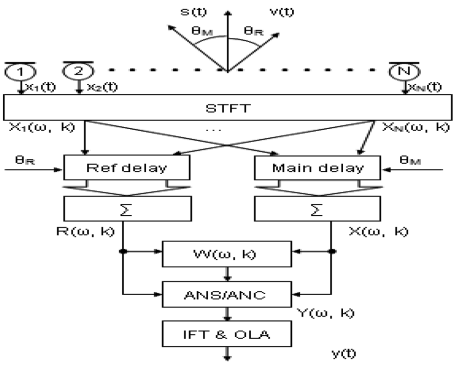

Fig. 1. The structure of the system

Consider the formation of the main signal and reference signal for the case of the target speaker and the noise source. The sound propagation model for each beam can be described as follows:

$$X(\omega,k) = S(\omega,k) + H_v(\omega,k)\mathrm{V}(\omega,k) = \mathrm{S}(\omega,k) + N_x(\omega,k), \tag{2}$$

$$R(\omega,k) = H_r(\omega,k)V(\omega,k) + H_s(\omega,k)S(\omega,k), \tag{3}$$

where $S(\omega,k)$ – is the signal of the target speaker, $\mathrm{V}(\omega,k)$ – is the signal of the noise source arriving to microphone array from the direction θ_v, $H_v(\omega,k)$ – is the transfer function from noise source to main beam of FBF, $H_r(\omega,k)$ - is the transfer function from noise source to reference beam of FBF, $H_s(\omega,k)$ - is the transfer function from target speaker to main beam of FBF.

Adaptive noise suppression algorithm is as follows. ANS algorithm based on an assessment of the amplitude spectrum interference $\widetilde{N_x}(\omega,k)$ in the main channel, which is calculated from the spectrum of the reference channel signal using the noise equalizer:

$$\tilde{N}_x(\omega,k) = W(\omega,k-1)|R(\omega,k)|, \tag{4}$$

where $W(\omega,k)$ – is the transfer function of equalizer, k – is the time frame index. Estimation of the transfer function is calculated as follows:

$$W(\omega,k) = \begin{cases} (1+\beta)\times W(\omega,k-1), |X(\omega,k)| > \tilde{N}_x(\omega,k) \\ (1-\alpha)\times W(\omega,k-1), |X(\omega,k)| \le \tilde{N}_x(\omega,k) \end{cases}, \tag{5}$$

where α is the release rate, β is the attack rate.

The algorithm in the absence of a speech signal $|\mathrm{S}(\omega,\ \mathrm{k})| \approx 0$ aligns the noise amplitude spectrum of the reference channel to the noise spectrum in the main channel:

$$|H_v(\omega,k)V(\omega,k)| \approx W(\omega,k)|H_r(\omega,k)V(\omega,k)|. \tag{6}$$

Since the spectrum of noise in the reference channel is usually much greater than the noise spectrum in the main channel: $|H_r(\omega,k)V(\omega,k)| \gg |H_v(\omega,k)V(\omega,k)|$, than $W(\omega,k) \ll 1$. Therefore it is necessary to constrain the maximum values of the transfer function of the equalizer:

$$\tilde{W}(\omega,k) = Min\{W_{max}(\omega), W(\omega,k)\}, \tag{7}$$

where $W_{max}(\omega)$ is the maximum values of the transfer function of the equalizer.

Constraint of the transfer equalizer function prevents unwanted amplification of the target spectral signal components belonging to a reference channel.

The estimation of the noise spectrum of the main channel is used in the algorithm of dual channel spectral subtraction to estimate spectrum of the target signal:

$$\left|\tilde{S}(\omega,k)\right| = |X(\omega,k)| - \tilde{N}x(\omega,k) = g(\omega,k) \times |X(\omega,k)|, \tag{8}$$

where $g(\omega,k)$ is the gain function of spectral subtraction:

$$g(\omega,k) = |\tilde{S}(\omega,k)|/|X(\omega,k)| = SNR(\omega,k)/(1+SNR(\omega,k)), \tag{9}$$

where $SNR(\omega,k) = \left|\tilde{S}(\omega,k)\right|/\tilde{N}x(\omega,k)$ are the spectral SNRs. To reduce the residual musical noise we made the following modification of gain function:

$$g(\omega,k) = C \times [SNR(\omega,k)]^2, \tag{10}$$

where $C = 1\ldots.5$ is the slope of gain function.

In its final form, taking into account constraints of minimum and maximum values the spectral gain is as follows:

$$G(\omega,k) = Min\left\{1, Max\left\{G_0(\omega,k), C \times [SNR(\omega,k)]^2\right\}\right\}, \tag{11}$$

where $G_0(\omega,k)$−suppression spectral floor. Estimation of target signal after noise reduction is as follows:

$$\tilde{S}(\omega,k) = X(\omega,k) \times G(\omega,k). \tag{12}$$

The enhanced signal ŝ(t) is calculated, using Invers Fourier Transform (IFT) and overlap and add (OLA) technique.

3 Simulation Results

The proposed ANS method is compared with the following methods: Fixed Beam-forming (FBF), constrained frequency domain GSC [1], frequency domain Null-Steering Beamformer (NSB) [5] and Adaptive Noise Canceller in time and frequency domain (ANC-T, ANC-F) [11]. The comparison has been done for linear microphone array with 11 microphones with a inter microphone spacing 3.5 cm.

To test the noise reduction performance of these methods, a computer program has been developed. The comparison has done using Noise Reduction (NR) и Speech Distortion (SD) and SNR improvement measure (SNRI).

Noise Reduction. To estimate NR the coherent broadband interference (white Gaussian noise) arriving to the microphone array from the angle $+45^\circ$ was used. It has been set no useful signal $S(\omega, k) = 0$. Main channel beam was steered in the look direction $\theta_M = 0^\circ$, the beam of the reference channel is steered in the direction $\theta_R = +45^\circ$:

$$\mathrm{X}(\omega, k) = H_v(\omega, k)V(\omega, k), \qquad \mathrm{R}(\omega, k) = H_r(\omega, k)V(\omega, k). \tag{13}$$

The NR was calculated using the residual interference signal power with interference power in a separate microphone:

$$NR\,dB = 10\log[P_{mic}/P_{out}]. \tag{14}$$

In this case, NR_{mic} on a separate microphone is equal to 0 dB [3].

Speech Distortion. To estimate SD the coherent speech signal arriving to the microphone array from the angle 0° was used. It has been set no interference signal $V(\omega, k) = 0$. Main channel beam was steered in the look direction $\theta_M = 0^\circ$, the beam from the reference channel is steered in the direction $\theta_R = +45^\circ$:

$$\mathrm{X}(\omega, k) = \mathrm{S}(\omega, k), \qquad \mathrm{R}(\omega, k) = H_s(\omega, k)S(\omega, k). \tag{15}$$

In this case, the output of microphone array FBF is undistorted speech signal:

$$Y_{FBF}(\omega, k) = S(\omega, k). \tag{16}$$

The speech signal is distorted if other methods are used. The ratio of the power of the distorted and undistorted signals y(t) and s(t) is defined as speech distortion:

$$SD\,dB = 10\log[P_S/P_Y] = 10\log[P_{FBF}/P_{out}] \tag{16}$$

SNR Improvement. To estimate SNRI coherent speech signal arriving to the microphone array from the angle of 0° and coherent broadband interference (white Gaussian noise) arriving from the angle of $+45^\circ$ were used. Main channel beam was steered in the look direction $\theta_M = 0^\circ$, the beam from the reference channel is steered in the direction $\theta_R = +45^\circ$. The input SNR has been set equal to −5 dB. SNRI evaluation was carried out in accordance with the procedure laid down in [4].

Table 1. NR, SD, SNRI after processing with different methods

Method	NR dB	SD dB	SNRI dB
FBF	10.98	0	5.65
NSB	30.16	7.97	16.86
GSC	14.80	0.11	10.03
ANC-T	45.43	11.78	9.85
ANS	**48.51**	**8.81**	**14.03**

The desired signal s(t) and interference n(t) are superposed with given SNR. The noisy signal x(t) is processed with the noise reduction algorithm. Afterwards the desired and interfering signals are separately processed with the resulting filter coefficients. SNRI was estimated by comparing outputs to inputs of the fixed filters.

The estimations of NR, SD, SNRI with different methods are shown in Table 1.

The proposed ANS method is superior to other methods according to the criterion of NR and close to NSB method for SD, SNRI criteria. Another advantage of the ANS is the ability to suppress diffuse noise. At the same time it is much inferior to GSC method for SD criteria. However, GSC loses its advantage under steering misadjustment, non-ideal microphones and reverberation multipath propagation.

4 Experimental Results in Real Conditions

4.1 Suppression of Partially Ccoherent Noise

We solved the problem of extracting speech speaker on the background of loud music using linear 8-microphone array with inter microphone spacing 5 cm.

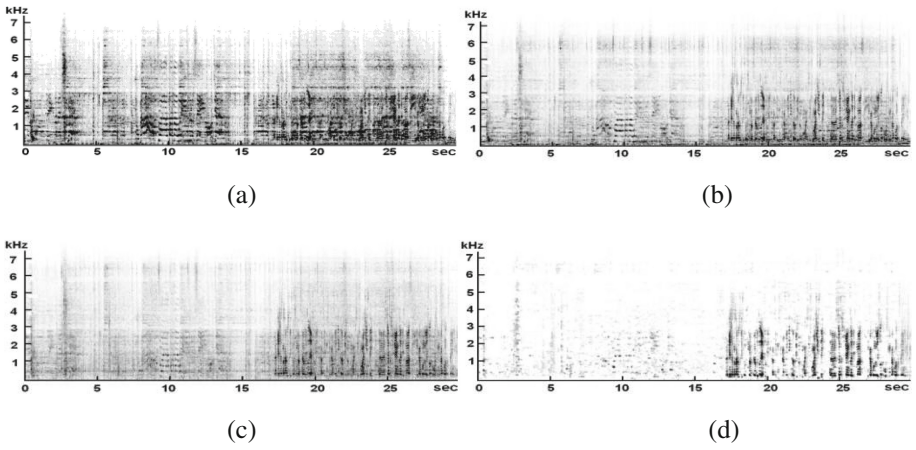

Fig. 2. Spectrograms for (a) Microphone signal, (b) FBF steered to target speaker, (c) FBF with ANC-T processing, (d) FBF with ANS processing.

Acoustic scenario: Office room size 6 × 13 × 3.2 m, reverberation time $T_{60} \approx 0.66$ s distance to the target speaker 3 m, $\theta_S \approx +10°$, distance to the loudspeaker 4.5 m, $\theta_V \approx -60°$, SNR ≈ −5.3 dB. Background music was a partially coherent, partially diffuse sound field. Background music is present throughout the range the target speaker's speech is present at 17–30 s interval.

The examples of the enhancement of speech with different methods are shown in Fig. 2.

The results of experiment are as follows. ANS gave the highest noise reduction comparing to the others: FBF (8 dB), FBF + ANC-T (11 dB), FBF + ANS (22 … 24 dB). ANS method showed the robustness to errors of microphone array steering on the target speaker and on the source of noise. ANS results reduction of both coherent and diffuse noise components.

4.2 Suppression of Diffuse Speech Interference

We solved the problem of the separation of two remote speakers speech in reverberant room using linear 6 × 8 microphone array with inter microphone spacing 5 cm. The

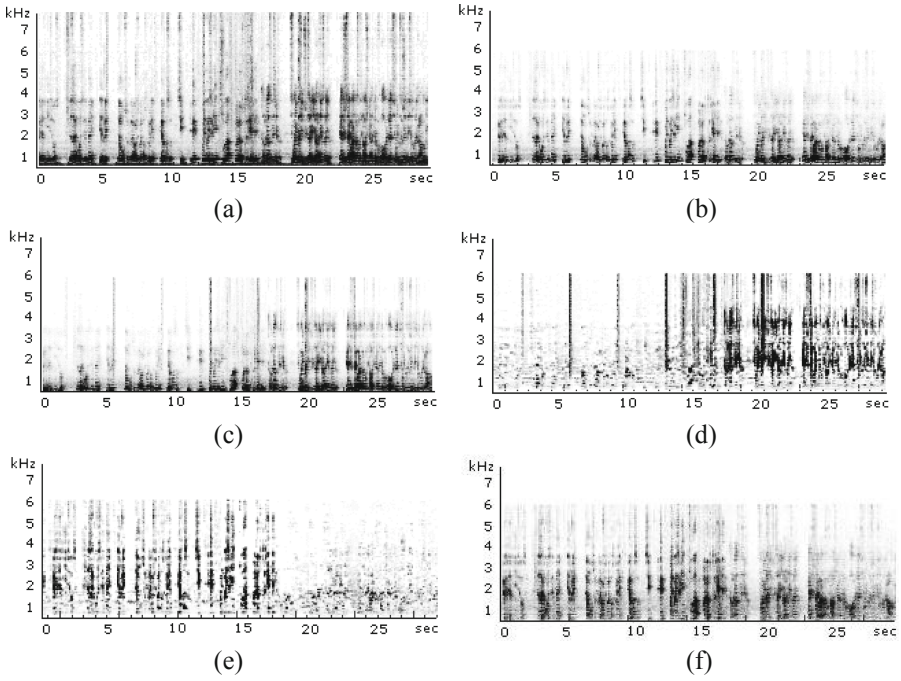

Fig. 3. Spectrograms for (a) microphone signal, (b) FBF steered to speaker 1, (c) FBF steered to speaker 2, (d) FBF with ANS enhancement of speaker 1, (e) FBF with ANS enhancement of speaker 2, (f) FBF with ANC-T enhancement of speaker 1.

low-pass filtering (6 kHz) for elimination of sidelobes was applied when processing signals in the microphone array.

Acoustic scenario: Office room size $6 \times 6.5 \times 3.2$ m, reverberation time $T_{60} \approx 0.88$ s, distance to the speaker_1 d1 = 6 m, $\theta_1 \approx +40°$, distance to the speaker_2 d2 = 5 m, $\theta_2 \approx 0°$. The speech was diffuse sound field. The speech of 1-st speaker is present on the time interval 0–15 s, the speech of 2-d speaker is present on the time interval 15–30 s. The examples of the enhancement of speech with different methods are shown in Fig. 3.

The results of experiment are as follows. ANS method allowed separate the remote speakers in reverberant room. Maximum suppression of the target speaker in using ANS is in the frequency range of 0–500 Hz, where the main lobe of array beampattern is broad and leakage of the target signal in the reference channel is the maximum. ANS method suppresses speech of interfering speaker significantly more effectively than the FBF, ANC-T, ANC-F methods.

5 Conclusion

A new speech enhancement method with MA for the joint suppression of coherent and diffuse noise is presented. The method is based on combined target and noise steering beamforming and algorithm of ANS. The ANS is based on the algorithm of dual channel spectral subtraction. The spectral subtraction results in the reduction of coherent and diffuse noise in the target beam signal. The proposed ANS yields better SNR improvement than conventional FBF and GSC algorithms and the best noise reduction comparing FBF, GSC, NSB algorithms and microphone alignment technique [12]. The experimental results show that ANS is an efficient procedure for speech enhancement in real life noisy and reverberant conditions with SNRs down to −5.3 dB and reverberation time up to $T_{60} \approx 0.88$ s. The additional advantages of ANS are its low computational cost that allows real-time speech processing and robustness to errors MA steering on the target speaker and source of noise.

Acknowledgements. This work was partially financially supported by the Government of the Russian Federation, Grant 074-U01.

References

1. Fischer, S., Simmer, K.: An adaptive microphone array for hands–free communication. In: Proceedings of IWAENC-1995, pp. 1–4 (1995)
2. McCowan, I.A.: Robust Speech Recognition using Microphone Arrays. Ph.D. Thesis, Queensland University of Technology, Australia (2001)
3. Brandstein, M., Ward, D. (eds.): Microphone Arrays. Springer, Heidelberg (2001)
4. Benesty, J., Makino, S., Chen, J. (eds.): Speech Enhancement. Springer, Heidelberg (2005)

5. Jonhson, D.H., Dungeon, D.E.: Array Signal Processing: Concepts and Techniques. Prentice-Hall, Upper Saddle River (1993)
6. Spalt, T., Fuller, C., Brooks, T., Humphreys, W.: A Background Noise Reduction Technique using Adaptive Noise Cancellation for Microphone Arrays. American Institute of Aeronautics and Astronautics, pp. 1–16 (2011)
7. Cao, Y., Sridharan, S., Moody, M.P.: Post-microphone-array speech enhancement with adaptive filters for forensic application. In: Proceedings of International Symposium on Speech, Image Processing and Neural Networks, pp.253–255 (1994)
8. Meyer, L., Sydow, C.: Noise cancelling for microphone arrays. In: Proceedings of ICASSP-1997, pp. 211–213 (1997)
9. Jingjing, T. et al.: The algorithm research of adaptive noise cancellation based on dual arrays and particle swarm algorithm. In: Proceedings of International Conference on Environmental Engineering and Technology Advances in Biomedical Engineering, vol. 8, pp. 106–111 (2012)
10. Nathwani, K., Hegde, R.: Joint adaptive beamforming and echo cancellation using a non reference anchor array framework. In: Proceedings of Asilomar 2012, pp.885–889 (2012)
11. Bitzer, J., Brandt, M.: Speech enhancement by adaptive noise cancellation: problems, algorithms, and limits. In: Proceedings of 39-th AES Conference, pp. 109–113 (2010)
12. Stolbov, M., Aleinik, S.: Speech enhancement with microphone array using frequency-domain alignment technique. In: Proceedings of 54-th AES Conference (2014)