A PSEUDO ADAPTIVE MICROPHONE ARRAY

Jianfeng Chen, Shue Louis, Hanwu Sun
Institute for Infocomm Research, 21 Heng Mui Keng Terrace, Singapore 119613
Email: jfchen@i2r.a-star.edu.sg

ABSTRACT

A pseudo adaptive microphone array scheme is presented in this paper for high quality speech acquisition. This two-stage scheme consists of an initial Enhanced Cross-Power Spectrum Phase (ECPSP) stage, which is used to estimate the directions of the target as well as the spatially distinguishable undesired signals, followed by an optimum spatial filter, constructed from the orthogonal components of the interferer subspace. The proposed method may thus be able to minimize the energy in the estimated directions corresponding to the interferers. Since the spatial filter is obtained solely from the estimated directions of the interferers, some difficulties encountered in adaptive beamforming technology, e.g., the negative impact of reverberation, stability issues in adaptive control in case of multiple or nonstationary interferers, target-signal canceling and distortion, can be avoided to some extent. Simulation results are provided to demonstrate the effectiveness of the proposed scheme.

1. INTRODUCTION

There are numerous situations where it is necessary to enhance the quality of speech signals which have to be captured in noisy environments. This includes teleconferencing, surveillance, hands-free talking, voice communications and so on. Microphone arrays have been shown to be capable of improving the quality of the incoming speech in recent years [1]-[7]. Several possibilities exist for the processing of broadband signals obtained from an array of microphones. Delay-and-sum beamformer [1][2] allows for a relatively simple processor, where a signal source from a desired direction is passed while sources originating from other directions are attenuated to some extent. However, delay-and-sum beamformers in general require a large number of microphones to achieve high performance, especially for the low frequency components. Adaptive beamformers [3]-[6], while more complex, can provide more attenuation to a given interfering source than is afforded by a comparable delay-and-sum beamformer. In this connection, the use of adaptive microphone beamformers has been extensively studied and the related research works are still active in recent years.

However, some difficulties still exist when using adaptive beamformers, examples being the negative impact of room reverberations [10], stability issues in the adaptive control schemes [11], potential cancellation and/or distortion of target-signal [5][6], etc. Since these problems arise from the use of adaptive algorithms, a possible resolution may be to depart from using such approach.

In this paper, a pseudo adaptive method is proposed which may outperform the adaptive approach in the cancellation of interferers, especially in practical situations where various interferers are present. The reason why it is described as 'pseudo adaptive' method lies in its unique working procedure. The scheme involves two stages: the directions of the multiple sources are estimated initially via the ECPSP method [9]. A spatial filter, constructed in frequency domain based on the result of the first stage, is subsequently employed for interferer cancellation. Depending on the nature of intended applications, the two stages are to be updated periodically, adaptively or manually.

The organization of this paper is as follows. A brief overview of the ECPSP method which we will adopt is given in Section 2. In Section 3 we will outline our proposed scheme for spatial filter design and interferer cancellation. Some practical considerations are summarized in Section 4, followed by the simulation results in Section 5. Some concluding remarks are provided in Section 6.

2. MULTIPLE SOURCE LOCALIZATION

In the simplest scenario, the direction of a single source can be obtained by estimating the time delay between two microphone outputs, which we will label as $x_i(n)$ and $x_j(n)$. The CPSP approach [9] has been used for this purpose because of its computational efficiency and stability. The cross-correlation coefficients of the two channels, CPSP$_{ij}(k)$, are defined as,

$$CPSP_{ij}(k) = \mathrm{DFT}^{-1}\left[\frac{\mathrm{DFT}[x_i(n)]\,\mathrm{DFT}^*[x_j(n)]}{|\,\mathrm{DFT}[x_i(n)]\,|\cdot|\,\mathrm{DFT}[x_j(n)]\,|}\right] \tag{1}$$

where DFT[·] (or DFT^{-1}[·]) denotes the Discrete Fourier Transform (or the Inverse Discrete Fourier Transform) and (*) is the complex conjugate operator, k is the time index of the window over which DFT/DFT^{-1} is carried out, and n the dummy variable. When there is only one source present, the time delay can be estimated by finding the maximum value of the CPSP$_{ij}(k)$, i.e.,

$$\tau = \arg\max_{k}[CPSP_{ij}(k)] \tag{2}$$

Then the source direction can be obtained by

$$\theta = \sin^{-1}\left[c\cdot\tau/(d\cdot F_s)\right] \tag{3}$$

where d is the distance between the two microphones, c the sound propagation velocity and F_s the sampling rate.

However, when there is more than one source, the estimation of the relative time delays is rather difficult due to the cross-correlation among different sources. For example, let s_1 and s_2 be the signals coming from two different directions. The outputs of the ith and jth microphones at time n are given by [10]

$$x_i(n) = a_i s_1(n+\alpha_i) + b_i s_2(n+\beta_i) \tag{4}$$

$$x_j(n) = a_j s_1(n+\alpha_j) + b_j s_2(n+\beta_j) \tag{5}$$

where α_i, α_j, β_i and β_j are the time delays, a_i, a_j, b_i and b_j are the distance attenuation coefficients, corresponding to s_1 and s_2, respectively. Substituting Eq. (4) and Eq. (5) into the numerator of Eq. (1), we have

$$\begin{aligned}\mathrm{DFT}[x_i(n)]\,\mathrm{DFT}^*[x_j(n)] &= a_i a_j S_1(\omega)^2 e^{-j\omega(\alpha_i-\alpha_j)} + b_i b_j S_2(\omega)^2 e^{-j\omega(\beta_i-\beta_j)}\\ &\quad + S_1(\omega)S_2(\omega)[a_i b_j e^{-j\omega(\alpha_i-\beta_j)} + a_j b_i e^{-j\omega(\beta_i-\alpha_j)}]\end{aligned} \tag{6}$$

Eq. (6) shows that CPSP method can be used to estimate the time delays when two signals are uncorrelated, i.e. the last term in Eq.(6) tends to zero. On the other hand, when the two signals are correlated, CPSP method fails to estimate the correct time delays. In fact, in this case CPSP$_{ij}(k)$ would show more peaks than expected: not only in correct directions but in incorrect directions due to the cross-correlation among different sources. Thus, in order to estimate the source directions reliably, it is necessary to suppress these undesired peaks. This problem can be resolved via a synchronous addition of the CPSP$_{ij}(k)$ derived from different microphone pairs.

It is known that if the positions of two microphones are changed, peaks of CPSP$_{ij}(k)$ corresponding to the 'correct' sources including desired and undesired signals and the 'fake' sources resulting from cross-correlation will change accordingly. Nevertheless, if microphone pairs are arranged so that they have a common acoustic center, peaks of CPSP$_{ij}(k)$ for the correct sources will be unchanged while peaks indicating resulting from cross-correlations will be different in general. Thus, under such conditions, if we add the CPSP$_{ij}(k)$ coefficients of all microphone pairs synchronously, the peaks corresponding to the correct source will be enhanced while the rest will not be in general.

Although the basic rule of microphone arrangement is clear, a well-designed array structure can nevertheless further facilitate the subsequent procedures. In [10] an equispaced array was adopted in their experiment. Such an arrangement is not in favor of subsequent synchronization of CPSP$_{ij}(k)$ of different microphone pairs and non-integer interpolation will be carried out, which would bring extra computation and error. To solve this problem, we adopted an array with unequal spacing between microphones. The distances of the microphone pairs are integer fraction of the largest distance (see Fig. 1). In this way, integer interpolation will be applicable to synchronize all CPSP$_{ij}(k)$.

3. OPTIMUM SPATIAL FILTER

In this section, we will discuss how the interferers are to be cancelled once their directions have been obtained. The scenario considered herein involves a single desired speaker signal s_1 and L-1 broadband point interferers s_l (l = 2,…,L). These sources are in the far field of a microphone array consisting of M omni-microphones (M >= L) arranged as shown in Figure 1. The array output vector in frequency domain can be written as

$$\mathbf{X}(\omega) = S_1(\omega)\mathbf{a}(\omega,\theta_1) + \sum_{l=2}^{L} S_l(\omega)\mathbf{a}(\omega,\theta_l) + \mathbf{N}(\omega) \qquad (7)$$

where $S_l(\omega)$ denotes the scalar discrete Fourier coefficient at frequency ω of l-th source s_l from direction θ_l, $\mathbf{a}(\omega,\theta_l)$ is an M×1 vector, denoting the array manifold vector for the same source, i.e., $\mathbf{a}(\omega,\theta_l) = \begin{bmatrix} 1 & e^{-j\omega d_2 \sin\theta_l / c} & \cdots & e^{-j\omega d_M \sin\theta_l / c} \end{bmatrix}$ with d_i (i = 2,…, M) indicating the distance between ith and first sensor. $\mathbf{N}(\omega)$ represents the background noise at frequency ω. In the following discussion, we assume the corrupting effect of $\mathbf{N}(\omega)$ is negligible relative to L-1 strong interferers, and hence focus simply on the interferers cancellation.

A spatial filter is a linear combiner which transforms the array data vector $\mathbf{x}$ into a scalar y via a weighting vector $\boldsymbol{w}$. In the frequency domain, this can be expressed as

$$Y(\omega) = \mathbf{w}(\omega)^H \mathbf{X}(\omega) \qquad (8)$$

where the superscript H denotes Hermitian transpose. Since the objective is to suppress the energy coming from directions $\theta_2, \ldots, \theta_L$, this motivates the following constraint for the selection of the weighting vectors $\mathbf{w}(\omega)$

$$\mathbf{w}(\omega)^H \mathbf{a}(\omega,\theta_l) = \begin{cases} 1 & for \;\; l = 1 \\ 0 & for \;\; l = 2,3,\cdots,L \end{cases} \qquad (9)$$

To satisfy the constraints of Eq. (9), $\mathbf{w}(\omega)$ must belong to the orthogonal component of the subspace spanned by the direction vectors corresponding to interferers s_l, l = 2,3, …, L. If we define the space matrix spanned by all the interferers as

$$\mathbf{A}_I(\omega) = \begin{bmatrix} \mathbf{a}(\omega,\theta_2) & \mathbf{a}(\omega,\theta_3) & \cdots & \mathbf{a}(\omega,\theta_L) \end{bmatrix} \qquad (10)$$

then we can prove the following weighting vector $\mathbf{w}(\omega)$ can meet the constraints of Eq. (9):

$$\mathbf{w}(\omega) = \frac{\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_1)}{\mathbf{a}^H(\omega,\theta_1)\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_1)} \qquad (11)$$

where $\mathbf{P}^{\perp}(\omega)$ is a minimum-norm solution to the null space of $\mathbf{A}_I(\omega)$ given by

$$\mathbf{P}^{\perp}(\omega) = \mathbf{I} - \mathbf{P}_I(\omega) = \mathbf{I} - \mathbf{A}_I(\omega)[\mathbf{A}_I^H(\omega)\mathbf{A}_I(\omega)]^{-1}\mathbf{A}_I^H(\omega) \qquad (12)$$

and $\mathbf{I}$ is the M×M unit matrix. It should be noted that the following relationship holds:

$$[\mathbf{P}^{\perp}(\omega)]^H = \mathbf{P}^{\perp}(\omega) \qquad (13)$$

The proof is straightforward. Substituting Eq.(11) into Eq. (9), with l =1, and considering the relationship in Eq. (13), we can directly obtain

$$\mathbf{w}(\omega)^H \mathbf{a}(\omega,\theta_1) = \frac{\mathbf{a}^H(\omega,\theta_1)\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_1)}{\mathbf{a}^H(\omega,\theta_1)\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_1)} = 1 \qquad (14)$$

When l =2,3,…,L, we can rewrite Eq. (9) as

$$\mathbf{w}(\omega)^H \mathbf{a}(\omega,\theta_l) = \frac{\mathbf{a}^H(\omega,\theta_1)\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_l)}{\mathbf{a}^H(\omega,\theta_1)\mathbf{P}^{\perp}(\omega)\mathbf{a}(\omega,\theta_1)} \qquad (15)$$

Since $\mathbf{P}^{\perp}(\omega)$ is the null space of $\mathbf{A}_I(\omega)$, we have $\mathbf{P}^{\perp}(\omega)\cdot\mathbf{a}(\omega,\theta_l) = \mathbf{0}$. Thus the numerator of Eq. (15) equals 0.

It should be mentioned that since there are no specific assumptions concerning the interferers or the relationship between the interferer and target-signal, the proposed method is appropriate in case of both stationary and nonstationary interferers, and the performance of this method is not linked to the correlation among the signal sources.

4. PRACTICAL CONSIDERATIONS

In practice, the number of sources L is an unknown quantity which needs to be determined online. Although the CPSP(k) coefficients can be enhanced by synchronous addition, it is still not easy to distinguish the true sources from the 'fake' sources in real applications. Generally an empirical threshold should be used in such case, but this would make the result over sensitive to the environment.

Here let us consider this problem from a pragmatic perspective. Since we make the assumption that the number of sources L is no more than the number of microphones M, we can simply choose the largest M peaks from CPSP(k) as estimates of the directions for the unknown number of sources. It is shown in our simulation that the performance of the proposed method would barely be affected as long as all the correct sources are included and regardless of the number of false sources. In fact, the proposed method can also deal with the case where L>M, but in this case only the first M strongest interferers will be cancelled while others will remain in the output. This feature further confers the proposed method the ability of dereverberation since most of reverberation energy is contained in the first few reflections.

In practice, considering the limited sensors and the array dimension, it is better to set L to be less than M so as to ensure the good performance in the first few strongest interferers. On the other hand, to avoid signal canceling, a 'buffer zone' around the target direction needs to be set to avoid generating notches near the target-signal and thereby causing signal distortion.

Within the framework of our proposed method, online target and interferer tacking can be implemented by updating the directions of the sound source duly, via some control logic mechanism. While accurately discriminating the desired sound from the interferences remains an open issue, anomalies can be minimized to an acceptable level with prior knowledge and use of control rules. As regards the estimation biases, the proposed method does possess certain robustness properties. The main negative consequence is that a given interferer may not be suppressed optimally, rather than suppressing by mistake the desired signal as can occur in adaptive methods. However, it needs to be pointed out that as long as an interferer is a sufficiently active and strong sound source, in practice it will most likely to be detected and cancelled during the operation period of the proposed method.

5. SIMULATION RESULTS

Computer simulations have been carried out using the omni-microphone array with eight sensors, as shown in Figure 1. The array size *d* was set to be 0.80 meter. Sampling frequency was 16 kHz. In the first situation, three highly correlated broadband Gaussian white noises impinged on the array from –20°, 0° and 40°, respectively. Figure 2(a)-2(d) illustrate the $\mathrm{CPSP}_{ij}(k)$ coefficients calculated from microphone pairs 1-8, 2-7, 3-6 and 4-5, respectively. Figure 2(e) shows the result of $\mathrm{CPSP}(k)$, synchronous additions of four $\mathrm{CPSP}_{ij}(k)$ coefficients. It can be seen that the peaks corresponding to the 'correct' sources have been emphasized while the peaks for the 'fake' sources are attenuated to a quite lower level.

In the second situation, the desired sound source, located in 0°, was a loudspeaker repeatedly playing a sentence randomly chosen from the TIMIT database. Two broadband Gaussian white noises (interferers) came from 25.3° and 40° respectively. The filter response of the spatial filter, constructed using Eq.11, is shown in Fig. 3. The two deep notches generated at the directions of the interferers indicate the strong cancellation capability. A comparison of the proposed algorithm with two classical beamformers, namely delay-and-sum beamformer (uniformly weighted) and Griffiths-Jim beamformer [4] (filter taps: 25), is shown in Fig.4. It can be seen that the proposed algorithm can significantly improve the quality of speech by eliminating the energy of interferers, resulting in a cleaner speech with low distortion. Furthermore the quantitative evaluation of the proposed method has been done by using signal-interference-ratio (SIR) as the measurement. Statistical analysis shows that in the scenario introduced above the proposed method can obtain SIR improvement up to 34 dB compared to when only a single microphone is used, and is also much higher than the other two beamformers (See Table 1).

Table 1: Comparison result of signal-to-interferer ratio

Method	SIR (dB)
Single microphone	-11
Delay-and-sum	-2
Griffiths-Jim	7
Proposed method	23

In the third example, the proposed method was used in a highly reverberant environment. Results herein described refer to a room of size 3m×4m×5m, with β=0.8 for the reflection coefficients of the wall, the ceiling and floor. The array was unchanged and was placed near one of the walls. The loudspeaker (the desired source) was located at three meters away from the center of the array with no interferer present. The processed signals of using the three methods are shown in Fig. 5. From both the waveform envelope and subjective listening test, it was verified that the proposed method is able to recover the original signal better than the other two beamformers in this case.

6. CONCLUSIONS

The method proposed in this paper can overcome most of difficulties of adaptive beamforming while still maintaining superior performance in interferer cancellation. Since no adaptation procedure is involved in the algorithm, the method is free from problems arising from the adaptive approach. Moreover, the proposed method still possesses the self-learning ability by way of strategically updating the directions of sound sources. The instance of updating can be controlled in automatic, manual or mixed mode. This provides user great flexibility in customizing the system, which may be crucial in taking the full advantage of the method. Finally, the proposed method demonstrates a certain degree of false-tolerant ability in that however accurate the interferer directions are estimated, the proposed method can keep the target signal undisturbed and meanwhile minimize those correctly estimated interferers. Thus the proposed method is very competent in various applications, such as teleconferencing, hand-free communication applications, speech recognition system, home automation system, etc.

REFERENCES

[1] J. Flanagan, J. Johnston, "Computer-steered microphone arrays for sound transduction in large rooms," *J. Acoust. Soc. Amer.*, 78(5), pp.1508-1518, 1985

[2] W. Kellermann, "A self-steering digital microphone array," *Proc. ICASSP'91*, pp. 3581-3584, 1991

[3] O. Frost, "An algorithm for linearly constrained adaptive array processing," *Proc. IEEE,* 60(8), pp. 926-935, 1972

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

[5] H. Cox, "Robust adaptive beamforming," *IEEE Trans. on ASSP*, vol. 35, no. 10, pp. 1365-1375, Oct. 1987

[6] O. Hoshuyama, A. Sugiyama, "A robust adaptive beamformer for microphone arrays with a blocking matrix using constrained adaptive filters," *IEEE Trans. on Signal Processing,* vol. 47, no. 10, pp. 2677-2684, 1999

[7] M. Agrawal, S. Prasad, "Optimum broadband beamforming for coherent broadband signals and interferences," *Signal Processing,* 77, pp.21-36, 1999

[8] M. Omologo, P. Svaizer, "Use of the crosspower spectrum phase in acoustic event location," *IEEE Trans. on Speech and Audio Processing,* vol. 5, no. 3, pp.288-292, May 1997

[9] T. Nishiura, T. Yamada, "Localization of multiple sound sources based on a CSP analysis with a microphone array," *Proc. ICASSP'2000,* Turkey, June, 2000

[10] J. Bitzer, K. U. Simmer, K.-D. Kammeyer, "Theoretical noise reduction limits of the generalized sidelobe canceller (GSC) for speech enhancement," *Proc. ICASSP' 99,* pp. 2965 –2968, Arizona, Mar. 1999

[11] O. Hoshuyama, B. Begasse, "A. Sugiyama, A. Hirano, A real time robust adaptive microphone array controlled by an SNR estimate," *Proc. ICASSP'98,* pp. 3605-8, Seattle, May, 1998

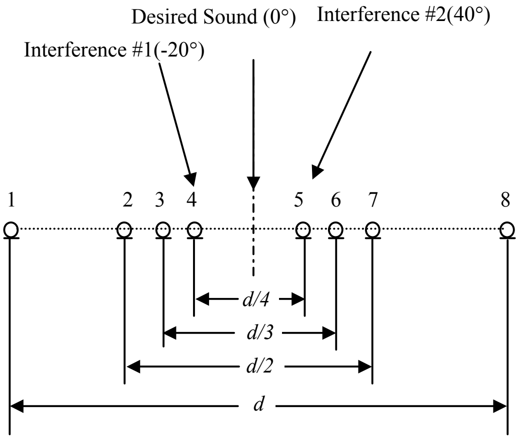

Figure 1: Diagram of the microphone array with eight sensors

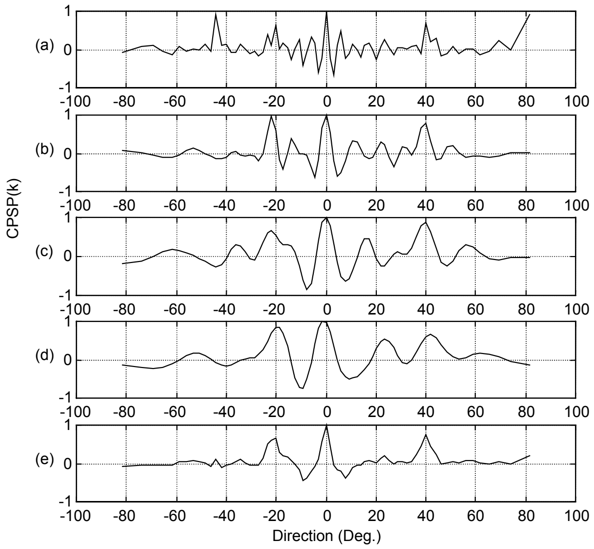

Fig. 2: CPSP coefficients of four microphone pairs (a) 1-8, (b) 2-7, (c) 3-6, (d) 4-5 and (e) their synchronous addition. Three sound sources are located at -20°, 0° and 40° respectively.

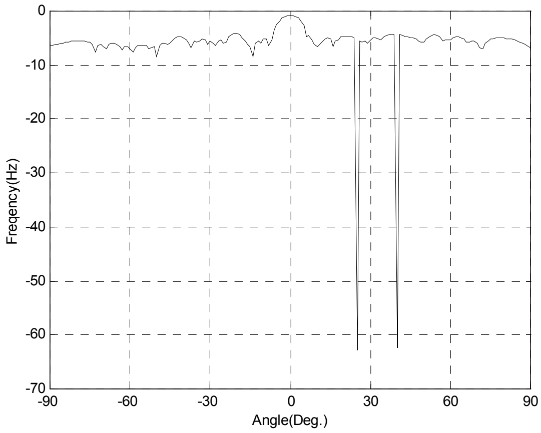

Fig. 3: Spatial filter response over frequency band from 500Hz to 4000Hz with two notches at 25.3° and 40°

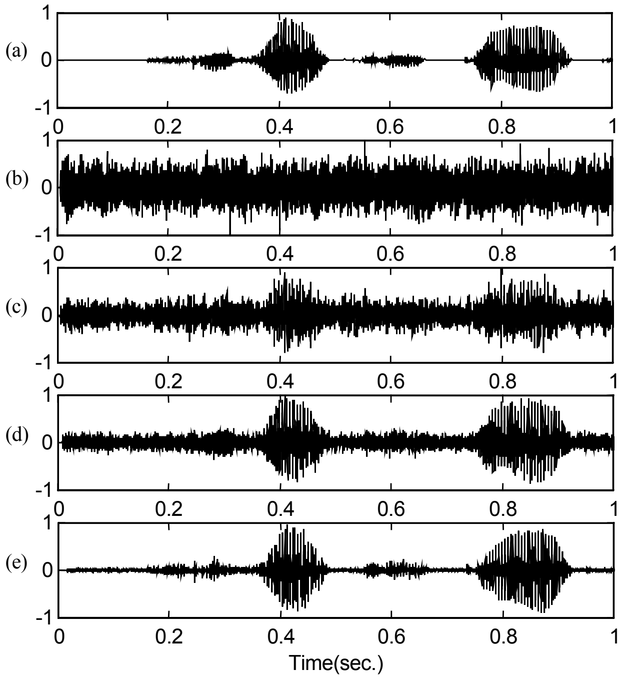

Fig. 4 Illustration of the improvement for the speech signal (corrupted by two broadband interferences located at 25.3° and 40°) using two kinds of classical beamformers and the proposed method. (a) - original speech, (b) – output of single microphone, (c) - output of delay-and-sum beamformer (uniformly weighted), (d) - output of Griffiths-Jim beamformer (25 taps), (e) - output of our proposed method

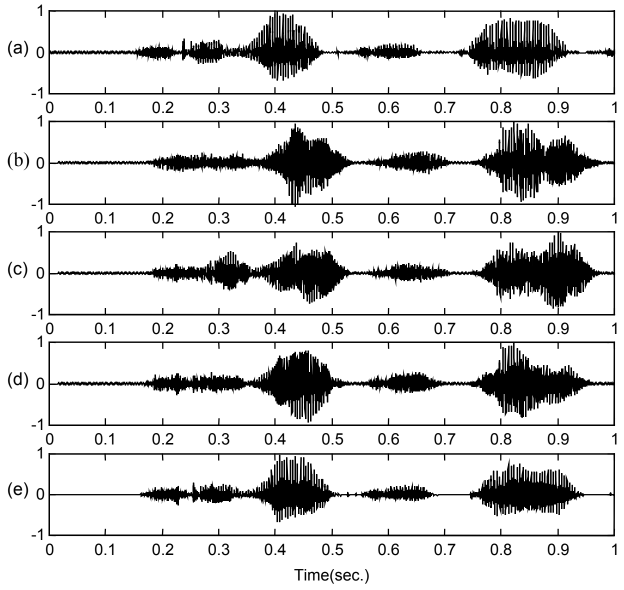

Fig. 5 Illustration of the improvement for the processed speech signal in reverberant environment using two kinds of classical beamformers and the proposed method. (a) - original speech, (b) – output of single microphone, (c) - output of delay-and-sum beamformer (uniformly weighted), (d) - output of Griffiths-Jim beamformer (25 taps), (e) - output of our proposed method