Journal of Applied Sciences

ISSN 1812-5654

science alert

ANSInet
an open access publisher
http://ansinet.com

Journal of Applied Sciences 13 (8): 1239-1244, 2013
ISSN 1812-5654 / DOI: 10.3923/jas.2013.1239.1244

Sound Localization and Directed Speech Enhancement in Digital Hearing Aid in Reverberation Environment

[1,2]Wang Qingyun, [2]Bao Yongqiang, [1]Zhao Li and [1]Meng Qiao
[1]Southeast University, Nanjing , Jiangsu, People's Republic of China
[2]Nanjing Institute of Technology, Nanjing, Jiangsu, People's Republic of China

Abstract: A new scheme of source localization and directed speech enhancement in digital hearing aid in reverberation environment is proposed in this paper. In the field of source localization, an adaptive PSP-AED time delay estimation algorithm is proposed. With the position of the source, the scheme enhances the speech signal using broadband Linearly Constrained Minimum Variance (LCMV) beamformer. The interferences and the noises are eliminated when the beampattern is directed to the target source position. The proposed algorithm is robust to noise. It estimates the time delay accurately when Signal to Noise Ratio (SNR) is lower than 0 dB and then the position of the source is acquired. The improvement of Signal to Interference Ratio (SIR) is achieved upper than 7 dB in reverberation environment. Numerical Experiments prove that the performance of wave quality, spectrogram and SIR is improved notably in the enhanced speech signal.

Key words: Digital hearing aid, source localization, directed speech enhancement

INTRODUCTION

The performance of digital hearing aid degrades with the decrease of Signal to Noise Ratio (SNR) in noisy and reverberation environment (Pandey and Mathews, 2011; Kamkar-Parsi and Bouchard, 2011). Regular de-noising algorithm is aim to eliminate the stationary noise (etc. white noise), which often fails when non-stationary noise exists, such as trumpet, whistle and thunder. Moreover, the interference speech also leads to the difficulty for target speech understanding.

Sound localization and directed speech enhancement are effective methods to increase SNR of target speech and improve the intelligibility of hearing loss people by digital hearing aid. In 2001, Bernard Windrow applied 6-elements liner array and fixed beam forming technique in digital hearing aid at first time and improved the intelligibility of a severe deafness patient significantly (Widrow, 2001). But the scheme of Bernard Windrow supposed that the target speech was always located in the front of the patient, rather than localized and tracked the speech in arbitrary direction. The scheme (Wu *et al*., 2007) researched 3D sound localization method, which calculated the azimuth and elevation angel by the time delay estimation of 4-elements square microphone array and then enhanced the target speech by adaptive beam forming. In addition, Head Related Transfer Function (HRTF) was proposed in recent years for sound localization (Keyrouz and Abou Saleh, 2007; Supper *et al*., 2006).

In the reverberation environment, sound wave received by microphone reflects by walls, ceiling board, furniture, people's head and shoulder, which reduce accuracy of localization algorithm. Based on the microphone signal processing and adaptive signal estimation, we propose a new sound localization and directed speech enhancement scheme in this study. In the scheme, PSP-AED time delay estimation is utilized to acquire the position of the target speech source and then broadband Linearly Constrained Minimum Variance (LCMV) beam is steered to the target speech to eliminate noises and interferences. Experiments and simulations for 4-elements glasses hearing aid demonstrate that the waveform, spectrogram and Signal to Interference Ratio (SIR) are improved and the pure target speech are recovered effectively.

SOUND LOCALIZATION AND DIRECTED SPEECH ENHANCEMENT SCHEME

The proposed localization and directed speech enhancement scheme is shown in Fig. 1. $y_1(k)...y_N(k)$ are signals received by N-elements microphone array. These signals are utilized by PSP-AED time delay estimation and sound localization algorithm to calculate the position of the target speech source. The enhanced output speech z(k) is then acquired through spatial-time filtering combined with the speech position. The coefficients of spatial-time filter H(z) are updated by broadband LCMV constraint optimization step.

Corresponding Author: Wang Qingyun, Southeast University, Nanjing, People's Republic of China

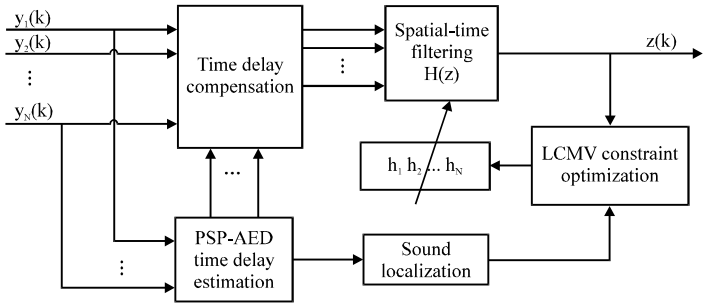

Fig. 1: Sound localization and directed speech enhancement scheme

PSP-AED time delay estimation algorithm: Adaptive Eigenvalue Decomposition algorithm (AED) algorithm is derived from room reverberation model (Benesty, 2000; Benesty *et al*., 2007). It is assumed that g_1 and g_2 are paths with length M from speech source s(k) to a pair of microphones in a room:

$$g_i = [g_{i,0} \quad g_{i,1} \quad \cdots \quad g_{i,M-1}]^T, i = 1,2 \tag{1}$$

Then signals received by microphones are:

$$y_i(k) = s(k) * g_i + v_i(k) = x_i(k) + v_i(k), i = 1,2 \tag{2}$$

Suppose $y_i(k)$, x(k) and $v_i(k)$ are vectors received by microphones, source speech and noise. If the noise $v_i(k)$ is ignored:

$$y_1^T g_2 = y_2^T g_1 \tag{3}$$

Construct:

$$\begin{aligned} y(k) &= [y_1^T(k) \quad y_2^T(k)]^T \\ w &= [g_2^T \quad -g_1^T]^T \end{aligned} \tag{4}$$

We get:

$$y^T(k) w = 0 \tag{5}$$

The vector w is composed by room pulse responses g_1 and g_2. The index of the maxima of the room pulse response g_i indicates the time delay of the direct path from source speech to the microphone. The expectation of Eq. 5 can be get by:

$$E\{y(k)y^T(k)\}w = R_{yy} w = 0_{2M\times 1} \tag{6}$$

where, w is the eigenvector corresponding to the eigenvalue 0. If g_1 and g_2 share no common zeros and the autocorrelation matrix $R_{ss} = E\{s(k)s^T(K)\}$ of the source signals is of full rank, such a two-channel acoustic system is blindly identifiable. In practice, noise always exists and W is found as the normalized eigenvector of R_{yy} corresponding to the smallest eigenvalue, that is (Chen *et al*., 2006):

$$\hat{w} = \arg\min_{w} w^T R_{yy} w \tag{7}$$

Equation 7 can be estimated by adaptive algorithms, such as Normalized Least Mean Square (NLMS) and Recursive Least Square (RLS) algorithms. But the convergence of NLMS is rather slow in the noisy environment and the computational complexity of RLS is huge. In this study, a newly PSP-AED (parallel sub gradient projection AED) algorithm is proposed, by which the eigenvector w is adaptively estimated with better convergence and lower computation than those of traditional algorithms.

Consider the real Hilbert space $(\mathcal{H}, \langle\cdot,\cdot\rangle)$ and define $d(\hat{w}_k, C_k) = \|\hat{w}_k - P_{C_k}(w)\|, \forall \hat{w}_k \in \mathcal{H}$ standing for the distance from $\hat{\mathbf{W}}_k$ to a closed convex set C_k. Provided that $(\hat{\mathbf{W}}_k)_{k\in N} \subset H$ is satisfied, then a sequence of adaptive filter coefficient vectors can be obtained iteratively as:

$$w_{k+1} = \hat{w}_k + \lambda_k \left(\sum_{l\in I_k} \eta_l^{(k)} P_{H_l^-(\hat{w}_k)}(\hat{w}_k) - \hat{w}_k \right) \tag{8}$$

where the relaxation parameters satisfy $\lambda_k \in [0, 2]$ (Yamada *et al*., 2002). Thus, the parallel subgradient projection $P_{Ck}(\hat{\mathbf{W}}_k)$ onto the convex set C_k is defined as (Yamada and Yukawa, 2002; Yukawa *et al*., 2007):

$$P_{H_l^-(\hat{w}_k)}(\hat{w}_k) := \begin{cases} \hat{w}_k & \hat{w}_k \in H_l^-(\hat{w}_k) \\ \hat{w}_k + \dfrac{-g_l^{(k)}(\hat{w}_k)}{\|\nabla g_l^{(k)}(\hat{w}_k)\|^2} \nabla g_l^{(k)}(\hat{w}_k) & \hat{w}_k \notin H_l^-(\hat{w}_k) \end{cases} \tag{9}$$

where, g_l is a convex function, $\Delta g_l(\hat{\mathbf{W}}_k)$ is the sub gradient at point $\hat{\mathbf{W}}_k$, $I_k = \{l^k_1, ..., l^k_q\} \subset N$, $q \in N$, and:

$$\eta_l^{(k)} > 0, \sum_{l \in I_k} \eta_l^{(k)} = 1$$

For PSP-AED algorithm, assume:

- $w = [g^T_2 - g^T_1]T$ is the room pulse response from source speaker to a pair of microphones
- $y(k) = [y^T_1(k) - y^T_2(k)]^T$ is the received signal by the pair of microphones
- $Y_k = [y_k, y_{k-1}, \ldots, y_{k-r+1}]^T \epsilon R^{M \times r}$ is the matrix of the input signal of the estimation, where, r is the step
- $e_k = y^T_k w_k / \|w_k\| = [e(k), e(k-1), \ldots, e(k-r+1)]$ is the normalized error vector

Define the convex set:

$$C_k(\rho) := \left\{ \hat{w} \in \mathcal{H} : \frac{\left\| Y_k^T \hat{w}_k \right\|^2}{\left\| \hat{w}_k \right\|^2} - \rho \le 0 \right\} \quad (10)$$

and the convex function:

$$g(\hat{w}) := \frac{\left\| Y_k^T \hat{w} \right\|^2}{\left\| \hat{w} \right\|^2} - \rho, \forall \hat{w} \in \mathcal{H} \quad (11)$$

Then the gradient operator of the convex function:

$$t := \nabla g(\hat{w}) = 2Y \frac{Y^T \hat{w}}{\left\| \hat{w} \right\|^2}, \forall \hat{w} \in \mathcal{H} \quad (12)$$

Equation 8 and 9 can be calculated literately by substituting Eq. 11 and 12.

Source localization and directed enhancement system in glasses digital hearing aid: The localization precision of the microphone array is decided by the dimension of the array and the number of the microphones, that is, the more microphones the array concludes, the more accurate resolution it reaches. In digital hearing aid, the dimension of the microphone array and the number of the microphones are limited by the application. In this study, a glasses digital hearing aid with a 4-components square array is used to evaluate the proposed algorithm. The locations of the microphones and the speech source are illustrated in Fig. 2.

Suppose that the microphone m_1 locates the original point of the 3-D space and the real coordinates of the speech source is $S = [x, y, z]^T \epsilon R^3$ which can be calculated by the radius r, the azimuth angle θ and the elevation angle ϕ.

According to Fig. 2, r, θ and ϕ can be estimated by the distance differences among microphone pairs:

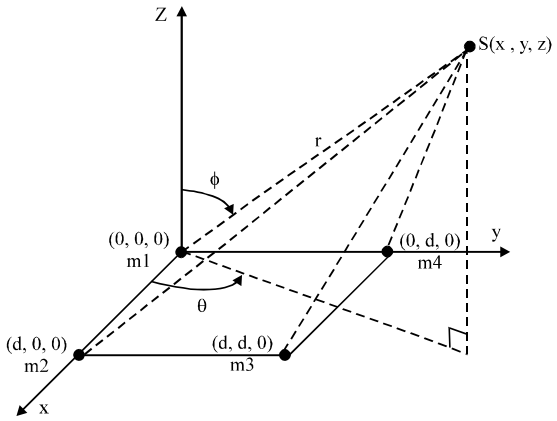

Fig. 2: Four components square microphone array

$$r = \sqrt{x^2 + y^2 + z^2} = \frac{d_{12}^2 + d_{14}^2 - d_{13}^2}{2(d_{12} + d_{14} - d_{13})} \quad (13)$$

$$x = \frac{d^2 - d_{12}(d_{12} - 2r)}{2d} \quad (14)$$

$$y = \frac{d^2 - d_{14}(d_{14} - 2r)}{2d} \quad (15)$$

$$\theta = \tan^{-1}\left[\frac{y}{x}\right] = \tan^{-1}\left[\frac{d^2 - d_{14}(d_{14} - 2r)}{d^2 - d_{12}(d_{12} - 2r)}\right] \quad (16)$$

$$\phi = \sin^{-1}\left[\frac{x}{r\cos\theta}\right] = \sin^{-1}\left[\frac{y}{r\sin\theta}\right] \quad (17)$$

Where:

$$d_{ij} = \frac{c\tau_{ij}}{f_s} \qquad i, j = 1, 2, 3, 4 \quad (18)$$

and τ_{ij} is the time delay between m_i and m_j which is depicted by samples and f_s is the sampling rate of the speech signals.

Broadband linearly constrained minimum variance (LCMV) beamforming: For broadband speech signal, the performance of Minimum Variance Distortion less signal Response (MVDR), Generalized Sidelobe Canceller (GSC) and narrow band Linearly Constrained Minimum Variance (LCMV) beamforming degrades (Benesty *et al.*, 2008; Saysar and Aldaihani, 2012). Broadband LCMV decomposes the speech signal into subbands by filterbank and processes LCMV beamforming in each subband and then the whole enhanced output signal is synthesized. The synthesized output signal:

$$z(k)=\sum_{n=1}^{N}h_n^T y_n(k) \quad (19)$$

where, h_n is the coefficient vector of the nth FIR filter. Substitute:

$$y_n(k)=g_n^T s(k), n=1,2,\cdots,N \quad (20)$$

into Eq. 19, the output:

$$z(k)=\left[\sum_{n=1}^{N}h_n^T G_n\right]s_L(k) \quad (21)$$

where, G_n is the room pulse response matrix from source speech to the nth microphone. In order to restore the source speech Distortion less, broadband LCMV filter is obtained by solving the following optimization problem (Li and Stoica, 2006):

$$\min_h h^T R_{yy} h, \text{subject to } G^T h = u$$
$$u=[1 \quad 0 \quad \cdots \quad 0]^T_{1\times L} \quad (22)$$

EXPERIMENTS AND SIMULATIONS

In all simulations and experiments the speech segments come from MULTIMIC speech database of Carnegie Mellon University, with the sampling frequency f_s = 16 kHz and the length 2 sec. The reverberant room has 2 meter width, 3 meter length and 3 meter high, with reverberant time T_{60} = 250 m sec. The locations of the microphone array and the speaker are demonstrated in Fig. 2. In order to decrease the computational complex we cut the room pulse response to the length 512 (Kuttruff, 2000). Simulations are repeated with different speech segments and the results are averaged.

Room pulse response estimation with different SNR: Adaptive Eigenvalue Decomposition (AED) algorithm estimates the room impulse response and finds the index of the maxima which corresponds to the time delay from the source speech to the microphone of the direct path. The estimations of the room impulse response h_2 (which indicates the path from the source speech to microphone m_2) by traditional NLMS-AED algorithm and the proposed PSP-AED algorithm with different SNRs are shown in Fig. 3, in which the left column shows the real response of h_2, the middle column shows the estimations of NLMS-AED algorithm with different SNRs and the right column shows the estimations of PSP-AED algorithm with different SNRs. The estimations of room impulse response from the source speech to other microphones achieve similar results.

Figure 3a shows when pure speech input without noise, both NLMS-AED and PSP-AED algorithms achieve good estimations of the index of the maxima of h_2 which indicates the time delay of the direct path. Figure 3b and c show when the SNRs of the input speech decrease, the estimation results of NLMS-AED algorithm deteriorate with obscured maxima of h_2, while PSP-AED algorithm holds the excellent estimation of time delay with cleared maxima of the h_2. Numerical experiments show that when SNR lower than 0 dB NLMS-AED algorithm fails, while such limit of PSP-AED algorithm is lower to -20 dB.

According to Eq. 13-18, r, θ and ϕ can be estimated by the distance differences among microphone pairs which is deduced by room pulse responses. Experiments demonstrate that when SNR = 0 dB the result (r, θ and ϕ) of PSP-AED algorithm is more accurate than that of NLMS-AED algorithm and GCC algorithm fails.

Directed speech enhancement: The goal of directed speech enhancement experiments is to improve the quality of the input speech segments. The spectrogram and Signal Interference Ration (SIR) of signals before and after enhancement are compared in the following. Spectrogram is a time-frequency analysis graphic of signal with time as abscissa and frequency as ordinate. The color or gray scale of the spectrogram indicates the intensity of the point of the signal. Twenty speech segments from MULTIMIC database are experimented. Figure 4a and b show the waveform and the spectrogram of target speech S_1 and Fig. 4c and d shows the waveform and the spectrogram of interference speech S_2.

The received signals of microphones m_1-m_4 are mixture of these two input signals in reverberation room. In audition test, the target speech S_1 cannot be identified and understood from the microphone received signal.

Figure 5 shows the waveform and spectrogram of enhanced output signal after the proposed broadband LCMV beam forming. Compared to the mixture signal of microphone, the waveform and the spectrogram of Fig. 5 are close to the target speech S0 and the audition test indicates better speech quality.

Minimum Variance Distortion less Response (MVDR), Generalized Sidelobe Canceller (GSC) and the proposed broadband LCMV beamforming experiments are conducted with plenty of speech segments input and Table 1 shows averaged enhancement results of SIR with different algorithms.

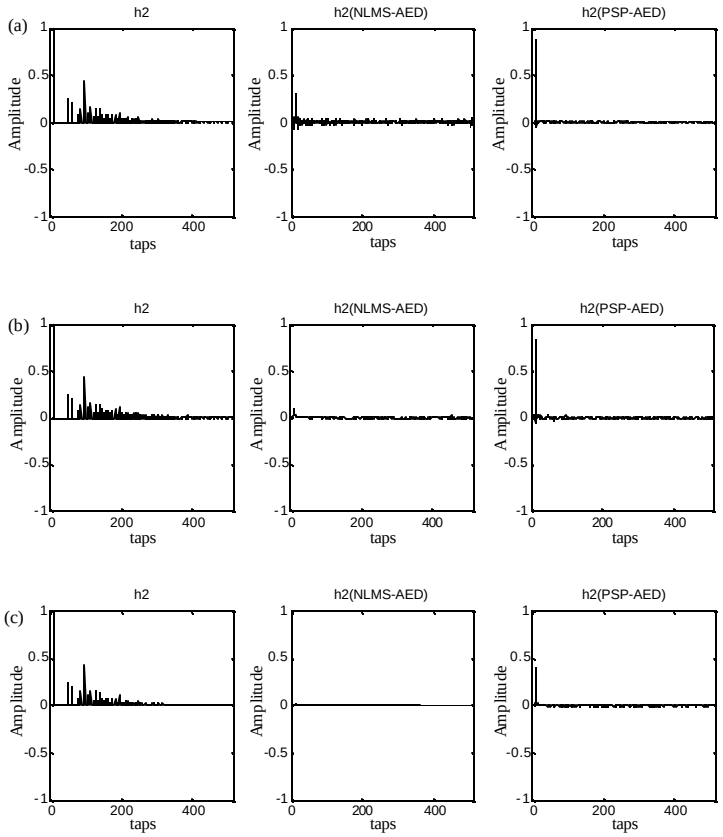

Fig. 3(a-c): Estimations of h_2, (a) Without noise, (b) SNR = 10dB and (c) SNR = 0dB

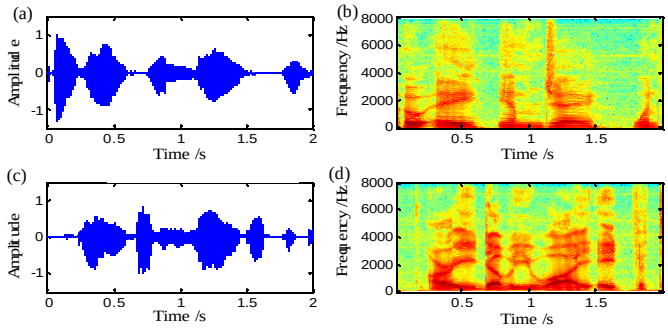

Fig. 4(a-b): (a) Waveforms of S_1 and S_2 and (b) Spectrograms of S_1 and S_2

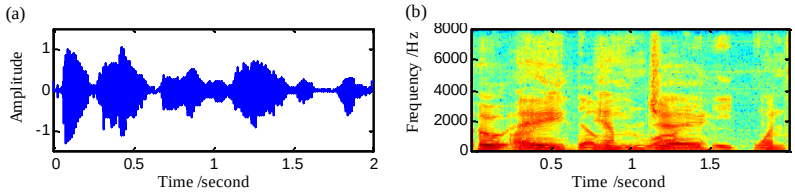

Fig. 5(a-b): (a) The output waveform of broadband LCMV and (b) Spectrogram of broadband LCMV

Table 1: The SIRs before and after enhancement

	Non-reverberation environment		Reverberation environment	
	SIR before (dB)	SIR after(dB)	SIR before(dB)	SIR after(dB)
MVDR	2.52	10.74	2.08	4.23
GSC	2.52	12.81	2.08	4.86
Broadband LCMV	2.52	13.04	2.08	7.45

Table 1 indicates that in non-reverberation environment, all of MVDR, GSC and the proposed broadband LCMV exhibit satisfying performance with 8-10 dB promotion of SIR. But in reverberation environment, the broadband LCMV beam forming achieves 5 dB promotion, better than MVDR (2.15 dB averaged) and GSC (2.78 dB averaged).

CONCLUSION

In this study a new 3-D localization and directed speech enhancement scheme for glasses digital hearing aids is proposed. Based on multichannel adaptive eigenvalue decomposition algorithm, the scheme estimates the impulse response coefficients from speech source to microphones by subgradient projection method, and then calculates the position of the speech source by geometric method. With the position of the source, the scheme enhances the speech signal in sub bands using wideband LCMV beam forming. The interferences and the noises are eliminated when the beam pattern is directed to the target source position. Compared with the traditional algorithms, the proposed algorithm is robust when reverberation and noise exist. Experiments and simulations demonstrate the validity of the scheme.

ACKNOWLEDGMENTS

This research was supported by China Postdoctoral Science Foundation (No. 2012M520973) and the Scientific Research Funds of Nanjing Institute of Technology (No. ZKJ201202). At last but not least, the authors would like to thank the reviewers for their valuable suggestions and comments.

REFERENCES

Benesty, J., 2000. Adaptive eigenvalue decomposition algorithm for passive acoustic source localization. J. Acoustical Soc. Am., 107: 384-391.

Benesty, J., J. Chen and Y. Huang, 2008. Microphone Array Signal Processing. Springer, Berlin, Germany.

Benesty, J., J. Chen, Y.A. Huang and J. Dmochowski, 2007. On microphone-array beamforming from a MIMO acoustic signal processing perspective. IEEE Trans. Audio, Speech Language Process., 15: 1053-1065.

Chen, J., J. Benesty and Y. Huang, 2006. Time delay estimation in room coustic environments: An overview. EURASIP J. Applied Signal Process., 10.1155/ASP/2006/26503

Kamkar-Parsi, A.H. and M. Bouchard, 2011. Instantaneous binaural target PSD estimation for hearing aid noise reduction in complex acoustic environments. IEEE Trans. Instrumen. Measur., 60: 1141-1151.

Keyrouz, F. and A. Abou Saleh, 2007. Intelligent sound source localization based on Head-related transfer function. Proceedings of the IEEE International Conference on Intelligent Computer Communication and Processing, September 6-8, 2007, Cluj-Napoca, pp: 97-104.

Kuttruff, H., 2000. Room Acoustic. Elsevier Science Publisher, New York.

Li, J. and P. Stoica, 2006. Robust Adaptive Beamforming. John Wiley and Sons, New York.

Pandey, A. and V.J. Mathews, 2011. Low-delay signal processing for digital hearing aids. IEEE Trans. Audio Speech Language Process., 19: 699-710.

Saysar, M. and M. Aldaihani, 2012. A stochastic model for analysis of manufacturing modules. Appled Math. Inform. Sci., 6: 587-600.

Supper, B., T. Brookes and F. Rumsey, 2006. An auditory onset detection algorithm for improved automatic source localization. IEEE Trans. Audio, Speech Language Process., 14: 1008-1017.

Widrow, B., 2001. A microphone array for hearing aid. IEEE Circuits Syst. Magazine, 1: 26-32.

Wu, W.C., C.H. Hsieh, H.C. Huang and O.T.C. Chen, 2007. Hearing aid system with 3D sound localization. Proceedings of the IEEE Region 10 Conference on TENCON 2007, October 30-November 2, 2007, Taipei, pp: 1-4.

Yamada, I. and M. Yukawa, 2002. An efficient robust stereophonic acoustic echo canceller based on parallel subgradient projection techniques. Singal Process., 1: 293-296.

Yamada, I., K. Slavakis and K. Yamada, 2002. An efficient robust adaptive filtering algorithm based on parallel subgradient projection techniques. IEEE Trans. Signal Process., 5: 1091-1101.

Yukawa, M., K. Slavakis and I. Yamada, 2007. Adaptive parallel quadratic-metric projection algorithms. IEEE Trans. Audio, Speech Language Process., 15: 1665-1680.