ACOUSTIC ECHO CANCELLATION FOR CONFERENCE SYSTEMS

Akihiko Sugiyama, Akihiro Hirano†*, and Kenji Nakayama*†

NEC Media and Information Research Laboratories
1-1, Miyazaki 4-chome, Miyamae-ku, Kawasaki 216-8555, Japan
† Dept. of Information and Systems Engineering, Kanazawa University
40-20, Kodatsuno 2-chome, Kanazawa 920-8667, Japan

ABSTRACT

An overview on acoustic echo cancellation for conference systems is presented. Design considerations for acoustic echo cancelers used in conference systems are first discussed. As one of the most challenging topics, multichannel acoustic echo cancellation is highlighted. The uniqueness problem is explained that is specific to multichannel echo cancellation, followed by solutions proposed to date.

1. INTRODUCTION

Echo cancellation is a widespread technology in communication systems [1]. An acoustic echo canceler [2] electrically models the path from the loudspeaker to the microphone by an adaptive filter. By exciting the adaptive filter with the signal from the remote side, or far-end, an echo replica is generated. The echo replica is subtracted from the microphone signal, resulting in an echo-canceled speech for transmission.

Acoustic echo cancelers are different in nature from those for line or data echo cancelers. The acoustic environment is open compared to that in the transmission line. Reflections on the walls generate a long impulse response for the echo path. Highly nonstationary back ground noise continuously changes the degree of disturbance in coefficient adaptation of the adaptive filter. Existence of the near-end (the local side) speech is another factor for interference in coefficient adaptation. Multichannel presentation of the far-end signal with multiple loudspeakers makes the configuration of the echo path more complicated. It is therefore worthwhile to discuss the overall design considerations for acoustic echo cancellation for conference systems.

2. DESIGN CONSIDERATIONS

There are a variety of design considerations in echo cancellation. Among these are residual echo level after convergence, convergence speed from the set-up, total computations, processing delay, noise immunity, and double-talk resistance to name a few. For the residual echo level, which defines the basic performance of echo cancellation, the number of taps is a dominant factor in relatively quiet environment*. It should be sufficiently large to cover a long impulse response of the echo path in the conference room. Moreover, the wider the signal bandwidth for better presence, the larger is the number of taps. Sometimes, it reaches several thousands and causes slow convergence with the popular NLMS (normalised least mean square) algorithm [3], and increased computations.

Convergence speed can be improved by fast convergence algorithms such as adaptive stepsize algorithms [4, 5] reflecting the progress of convergence, and for colored signals, Affine projection algorithms [6] or RLS (recursive least squares) algorithms [7]. From computational and stability viewpoints, the former two are common in acoustic echo cancellation. Subband echo cancelers [8] are useful for reducing the total computaions. Thanks to decimation after frequency-band division with a filterbank, the total computations are reduced in proportional to the number of subbbands. Aliasing and additional delay, which are introduced by the filterbank, can be reduced to a sufficiently small level by an oversampling filterbank [9] and a delayless subbband structure [10]. Further reduction in total computations is possible by adaptive intersubband tap assignment [11]. For noise immunity, it is effective to control the stepsize based on the signal-to-noise ratio [12, 13]. Double-talk, when the far-end and the near-end speech simultaneously exist, may cause degradation in echo cancellation because of the interference by the near-end speech, which is sometimes fatal. Therefore, coefficient adaptation should be disabled during double-talk periods. Double-talk detection plays a key role and is carried out based on coherence [14] or correlation [15, 16]. More than one measure can be combined for better detection [17].

A specific design consideration to acoustic echo cancelers is the number of channels for far-end signal presentation. With growing demands for enhanced presence in conference, use of multiple loudspeakers for presentation of the far-end speech has been spotlighted in these years. With multiple participants at different locations, multichannel presentation is helpful for speaker localization, resulting in easier speaker discrimination [18]. For multichannel conference systems, multichannel acoustic echo cancelers are needed.

With the structure based on linear combination [19], the echo paths are not correctly identified [20] for strongly cross-correlated input signals, like stereo speech signals. Actually, each filter coefficient is likely to misconverge to a final value which depends on the acoustic environment in the remote room [21]. Thus, any acoustic change in the remote room seriously degrades ERLE (echo return loss enhancement), which is the most common measure for echo cancelers. This is called the uniqueness problem [22].

3. MULTICHANNEL ECHO CANCELLATION

3.1 The Uniqueness Problem

Figure 1 depicts a traditional stereo echo canceler based on linear combination. The symmetry allows to consider only the echo $y_L(k)$, received at the left microphone. The time invariant $N \times 1$ coefficient vectors, $\mathbf{w}_{CL}(k)$, of the adaptive filters, the $N\times 1$ echo-path impulse response vectors, $\mathbf{h}_{CL}(k)$, the $M \times 1$ impulse response vectors, $\mathbf{g}_C$, of the acoustic paths in the remote room, the input signal vectors, $\mathbf{x}_C(k)$ and $\mathbf{x}(k)$, are defined by the following equations, where C stands for either L or R (**L**eft or **R**ight channel). k is the time index for discrete signals and $\mathbf{0}_{N\times 1}$ is a $N \times 1$ zero vector. Subscript CL denotes that the path is from the C channel to the L.

$$\mathbf{w}_{CL}(k) = [w_{CL,0}(k)\ w_{CL,1}(k)\ \cdots\ w_{CL,N-1}(k)]^T \quad (1)$$

$$\mathbf{h}_{CL} = [h_{CL,0}\ h_{CL,1}\ \cdots\ h_{CL,N-1}]^T \quad (2)$$

$$\mathbf{g}_C = [g_{C,0}\ g_{C,1}\ \cdots\ g_{C,M-1}]^T \quad (3)$$

*An FIR (finite impulse response) filter is naturally assumed here because, in practice, it is most widely used for echo cancellation.

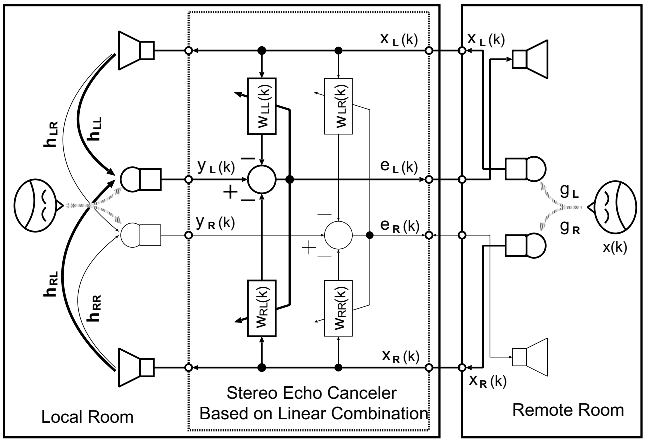

Figure 1: Stereo echo canceler based on linear combination.

$$\mathbf{x}_C(k) = [x_C(k)\ x_C(k-1)\ \cdots\ x_C(k-M+1)]^T \quad (4)$$

$$\mathbf{x}(k) = [x(k)\ x(k-1)\ \cdots\ x(k-M+1)]^T \quad (5)$$

The left-channel echo, $y_L(k)$ is given by

$$y_L(k) = \mathbf{h}_{LL}^T \mathbf{x}_L(k) + \mathbf{h}_{RL}^T \mathbf{x}_R(k). \quad (6)$$

The residual echo, $e_L(k)$, is obtained by subtracting the echo replicas which are related to the left input signal $\mathbf{x}_L(k)$ and the right input signal $\mathbf{x}_R(k)$, from the echo as

$$\begin{aligned} e_L(k) &= [\mathbf{h}_{LL} - \mathbf{w}_{LL}(k)]^T \mathbf{x}_L(k) \\ &\quad + [\mathbf{h}_{RL} - \mathbf{w}_{RL}(k)]^T \mathbf{x}_R(k). \end{aligned} \quad (7)$$

Assuming a single talker, the input signals, $\mathbf{x}_C(k)$ with $C = L$ and R are expressed by the convolution of the source signal vector $\mathbf{x}(k)$ with $\mathbf{g}_C$ as follows:

$$\begin{aligned} \mathbf{x}_C(k) &= [x_C(k)\ x_C(k-1)\ \cdots\ x_C(k-N+1)]^T \\ &= [\mathbf{g}_C^T \mathbf{x}(k)\ \mathbf{g}_C^T \mathbf{x}(k-1)\ \cdots\ \mathbf{g}_C^T \mathbf{x}(k-N+1)]^T \\ &= \mathbf{G}_C\ \tilde{\mathbf{x}}(k) \end{aligned} \quad (8)$$

The impulse response matrices in the remote room, $\mathbf{G}_C$ with $C = L$ and R, and an extended input signal vector $\tilde{\mathbf{x}}(k)$ are defined by (9) and (10).

$$\mathbf{G}_C = \left.\overbrace{\begin{bmatrix} \mathbf{g}_C^T & 0 & \cdots & 0 & 0 \\ 0 & \mathbf{g}_C^T & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & 0 & \cdots & 0 & \mathbf{g}_C^T \end{bmatrix}}^{N+M-1}\right\} N \quad (9)$$

$$\tilde{\mathbf{x}}(k) = [x(k)\ x(k-1)\ \cdots\ x(k-M-N+2)]^T \quad (10)$$

Substituting (8) for $C = L$ and $C = R$ in (7) gives

$$\begin{aligned} e_L(k) &= [\mathbf{h}_{LL} - \mathbf{w}_{LL}(k)]^T\ \mathbf{G}_L\ \tilde{\mathbf{x}}(k) \\ &\quad + [\mathbf{h}_{RL} - \mathbf{w}_{RL}(k)]^T\ \mathbf{G}_R\ \tilde{\mathbf{x}}(k). \end{aligned} \quad (11)$$

From (11), the mean-square error (MSE) $\mathrm{E}\{e_L^2(k)\}$ is obtained as

$$\begin{aligned} \mathrm{E}\{e_L^2(k)\} &= \mathrm{E}\{\mathbf{V}^T \tilde{\mathbf{x}}(k) \tilde{\mathbf{x}}^T(k) \mathbf{V}\} \\ &= \mathbf{V}^T \mathrm{E}\{\tilde{\mathbf{x}}(k) \tilde{\mathbf{x}}^T(k)\} \mathbf{V}, \end{aligned} \quad (12)$$

where $\mathrm{E}\{\cdot\}$ denotes a mathematical expectation operator and

$$\mathbf{V}^T = [\mathbf{h}_{LL} - \mathbf{w}_{LL}(k)]^T\ \mathbf{G}_L + [\mathbf{h}_{RL} - \mathbf{w}_{RL}(k)]^T\ \mathbf{G}_R. \quad (13)$$

In search of the minimum MSE, let us calculate the gradient ∇ of the MSE with respect to $\mathbf{V}$. From (12),

$$\begin{aligned} \nabla &= \frac{\partial \mathrm{E}\{e_L^2(k)\}}{\partial \mathbf{V}} \\ &= 2\mathrm{E}\{\tilde{\mathbf{x}}(k) \tilde{\mathbf{x}}^T(k)\} \mathbf{V}. \end{aligned} \quad (14)$$

The condition under which the minimum MSE is obtained, *i.e.* ∇ equals a null vector, leads to the following equation.

$$\begin{aligned} &\left\{[\mathbf{h}_{LL} - \mathbf{w}_{LL}(k)]^T \mathbf{G}_L + [\mathbf{h}_{RL} - \mathbf{w}_{RL}(k)]^T \mathbf{G}_R\right\} \\ &\qquad \cdot \mathrm{E}\{\tilde{\mathbf{x}}(k) \tilde{\mathbf{x}}^T(k)\} = \mathbf{0}_{1\times(N+M-1)}, \end{aligned} \quad (15)$$

where $\mathbf{0}_{1\times(N+M-1)}$ is the $1\times(N+M-1)$ zero matrix. Assuming that $\mathrm{E}\{\tilde{\mathbf{x}}(k)\tilde{\mathbf{x}}^T(k)\}$ has its inverse and postmultiplying both sides of (15) by $\mathrm{E}\{\tilde{\mathbf{x}}(k)\tilde{\mathbf{x}}^T(k)\}^{-1}$, it becomes

$$[\mathbf{h}_{LL} - \mathbf{w}_{LL}(k)]^T \mathbf{G}_L + [\mathbf{h}_{RL} - \mathbf{w}_{RL}(k)]^T \mathbf{G}_R = \mathbf{0}_{1\times(N+M-1)}, \quad (16)$$

where $\{\cdot\}^{-1}$ represents the inverse matrix. There is an infinite number of solutions to (16). Therefore, it does not imply that $\mathbf{w}_{LL}(k) = \mathbf{h}_{LL}$ nor $\mathbf{w}_{RL}(k) = \mathbf{h}_{RL}$. Correct echo-path identification is not achieved.

3.2 Solutions to the Uniqueness Problem

Because the uniqueness problem is caused by interchannel correlation in the far-end signals, the solution should naturally be decorrelation. However, in practical situations, the input signals have low-level uncorrelated components [23] and slight variations in the interchannel correlation [24]. Therefore, there are two general solutions; use of existing uncorrelated components and decorrelation.

3.2.1 Use of existing uncorrelated components

Uncorrelated components of cross-correlated input signals [23], or the variations in interchannel correlation [24] may be utilized for the unique coefficient vectors. It is also possible to enhance these effects by injecting uncorrelated noise components which are psychoacoustically masked by the original signal [25, 26]. A simplified masking technique assuming speech input is employed in [27]. Due to the impulse response tail effect [28], artificial tap-shortage can be utilized to generate uncorrelated noise components that serves as additive noise. Alternating coefficient adaptation for partitioning blocks of the adaptive filter implements this tap-shortage [29].

The correct echo-path identification can be achieved providing that the adaptation algorithm makes use of these uncorrelated low-level components or small variations in the interchannel correlation. However, convergence characteristics depend on the levels of the uncorrelated components, or the degrees of the variation in interchannel correlation. Small levels or small degrees of the variation require a fast-convergence algorithm with much computation and result in implementation difficulty.

3.2.2 Decorrelation

A typical decorrelation is preprocessing of the far-end signal. Among those are nonlinear processing [28, 30, 31, 32], FIR filtering with time-varying coefficients [33, 34], and IIR (infinite impulse response) filtering [35, 36, 37].

Partial decorrelation of input signals by introducing nonlinearity in each channel [28, 30] is now considered as a standard technique. It is also possible to adaptively control the degrees of nonlinearity based on the coherence of the far-end signals [31]. Several types of nonlinearities were compared from viewpoints of coherence, subjective quality, and sound localization [32]. The half-wave rectification provides

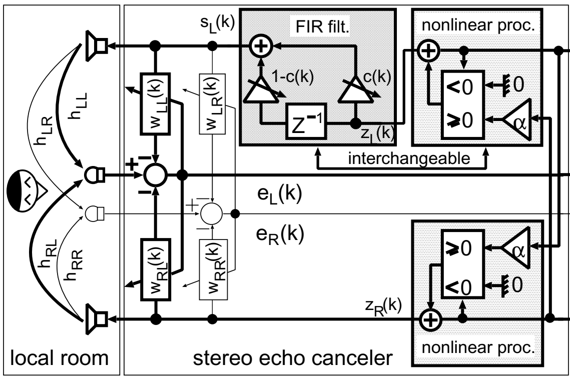

Figure 2: Combination of FIR filtering and twisted nonlinear processing.

the best quality with the simplest implementation. However, it is necessary to select different parameter values for speech and music. A smoothed rectification, obtained by paying the cost for standard-deviation estimation, does not need any adjustment while preserving as good quality as that with the half-wave rectification. Despite this decorrelation, convergence characteristics are still sensitive to the cross-correlation of the original input signals. The convergence of filter coefficients is slow unless a fast-convergence algorithm is used.

To obtain a satisfactory convergence speed with the NLMS algorithm, FIR filtering of the far-end signal with time-varying coefficients is effective [33, 34]. The NLMS algorithm requires less computations, making the implementation of a stereo echo canceler with FIR filtering easier. It introduces slides in the far-end signal for decorrelation, resulting in movement of the stereo image. This stereo image degradation is not significant according to subjective test results [34]. Convergence may be speeded up by more frequent slides. However, an optimum value in [34] exhibits a proven good compromise between convergence and speech quality.

IIR filtering [35, 36, 37] makes it possible to decorrelate the far-end signals with much smaller number of coefficients than that for FIR filtering. A desirable characteristic of IIR filtering was investigated through analysis [36]. A 2nd-order allpass filtering was shown to converge faster than its 1st-order counterpart [37]. However, there is no support or evaluation for subjective quality. This is the biggest uncertainty of IIR filtering at this moment.

3.2.3 Combination of Decorrelation Techniques

Multiple decorrelation techniques may be combined as far as the resulting convergence and speech quality are not degraded. A stereo echo canceler with FIR filtering and twisted nonlinear processing was proposed [38] to speed up convergence while preserving the subjective quality of the far-end signal. Its blockdiagram is depicted in Fig. 2. As is clear from the figure, it is not a simple combination of FIR filtering [34] and nonlinear processing [28]. The auxiliary signal to generate the nonlinear component is taken from another channel instead of the own channel, resulting in a twisted structure. Thus, the stereo input signals are partially mixed.

The FIR filtering generates interaural differences whereas the nonlinear processing emphasizes the inherent difference in the stereo signals. Therefore, the combination of the FIR filtering with nonlinear processing can be used to obtain faster convergence. It should be noted that the positions of FIR filtering and nonlinear processing are interchangeable.

The input-output relationship for the nonlinear transformation is

$$z_L(k) = \begin{cases} x_L(k) & \text{if } x_L(k) \geq 0 \\ x_L(k) + \alpha x_R(k) & \text{if } x_L(k) < 0 \end{cases}, \qquad (17)$$

for the left channel. $z_R(k)$ can be obtained in a similar manner. Mixing the stereo signals promotes the interchannel decorrelation, and therefore, it should speed up the convergence.

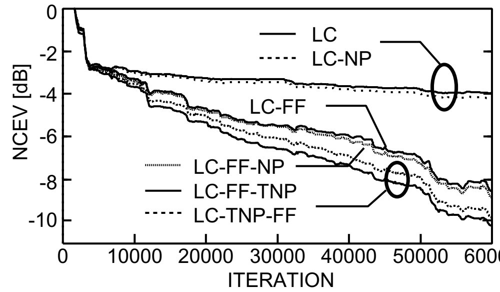

Figure 3: Simulation results for the proposed structure

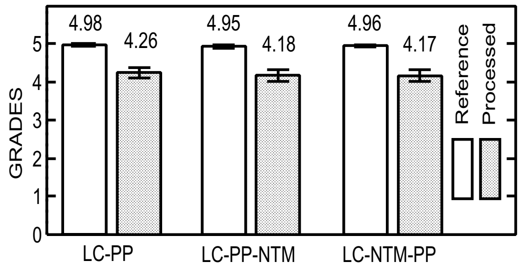

Figure 4: Subjective quality by listening test results.

Figure 3 shows convergence characteristics by the norm of the coefficient-error vector (NCEV) [34] for different structures with 1000-tap adaptive filters, -40dB background noise, and a recorded speech. It is clear that the structure LC-TNP-FF† converges as much as 20% faster than LC-FF‡. The exchanged placement of TNP and FF (LC-FF-TNP) provides comparable convergence characteristics to LC-TNP-FF. It should be noted that a simple combination of FIR filtering and nonlinear processing (LC-FF-NP) does not bring any additional effect compared to LC-FF. The twisted sturucture plays an important role.

Figure 4 shows the result of subjective assesment for the preprocessed signals. Six different speech and music signals were presented to 20 listeners based on the "triple stimulus/hidden reference/double blind approach [39]" for a score defined in the ITU-R five-grade impairment scale§ [40].

The top extremity of the vertical bar represents the mean value of the absolute grade, which is also numerically preented above the bar. The vertical line centered at the mean is the 95% confidence interval using a normal distribution. The audio quality of the processed signals does not show any statistical difference for all the structures under test. The impairment between the processed version and the reference is smaller than 0.8. This impairment is satisfactory for teleconferencing since the absolute grades are above 4.0.

3.3 Searching for a new solution

Use of uncorrelated components does not provide sufficient decorrelation, necessitating a fast convergence algorithm with higher complexity. Decorrelation could provide fast convergence with possible distortion and/or movement of the stereo image. There is a trade-off between the convergence

†Linear combination (LC) with twisted nonlinear processing(TNP) followed by FIR filtering (FF)

‡Linear combination with FIR filtering

§More detailed conditions are available in [38].

speed and the subjective quality of the far-end signal. Assessment of the far-end signal quality [32, 33, 34, 38] is essential when a decorrelation technique is applied. It is possible that a new solution with faster convergence and better far-end signal quality will be developed.

4. CONCLUSION

An overview on acoustic echo cancellation for conference systems has been presented with an emphasis on multichannel presentation. The uniqueness problem has been reviewed with solutions proposed to date. A good compromise between convergence and subjective quality will continue to be searched for.

REFERENCES

[1] Adaptive Echo cancellation for Speech Signals," Chap. 11, Advances in Speech Signal Processing, Ed. S. Furui and M. M. Sondhi, Marcel-Dekker, 1991.

[2] C. Breining, P. Dreiseitel, E. Hänsler, A. Mader, B. Nitsch, H. Puder, T. Schertler, G. Schmidt, and J. Telp, "Acoustic Echo Control," IEEE Signal Processing Mag., pp. 42-69, Jul. 1999.

[3] S. Haykin, "Adaptive Filter Theory, Third Edition," Prentice-Hall, 1996.

[4] R. W. Harris, D. M. Chabries, and F. A. Bishop, "A Variable Step (VS) Adaptive Filter Algorithm," IEEE Trans. ASSP-34, pp.309-316, Apr. 1986.

[5] V. J. Mathews and Z. Xie, "A stochastic gradient adaptive filter with gradient adaptive step size," IEEE Trans. SP, Vol.41, No.6, pp.2075-2087, Jun. 1993.

[6] K. Ozeki and T. Umeda, "An Adaptive Filtering Algorithm Using an Orthogonal Projection to an Affine Suspace and Its Properties," Elec. and Commun. in Japan, vol.67-A, pp.19-27, May 1984.

[7] J. Chioffi and T. Kailath, "Fast Recursive LS Transversal Filters for Adaptive Processing," IEEE Trans. ASSP-32, pp.304-337, Apr. 1984.

[8] A. Gilloire, "Experiments with Sub-band Acoustic Echo Cancellers for Teleconferencing," Proc. ICASSP'87, pp.2141-2144, Apr. 1987.

[9] W. Kellermann, "Analysis and Design of Multirate Systems for Cancellation of Acoustic Echoes," Proc. ICASSP'88, pp.2570-2573, Apr. 1988.

[10] D. Morgan and J. Thi, "A Delayless Subband Adaptive Filter Architecture," IEEE Trans. ASSP-43, No.8, pp.1819-1830, Aug. 1995.

[11] A. Sugiyama and F. Landais, " An Adaptive Intersubband Tap-Assignment Algorithm for Subband Adaptive Filters," Proc. ICASSP'95, pp.3051-3054, May 1995.

[12] S. Yamamoto and S. Kitayama, "An Adaptive Echo Canceller with Variable Step Gain Method," Trans. IEICEJ, Vol.E65, No.1, pp.1-8, Jan. 1982.

[13] A. Hirano and A. Sugiyama, "A Noise-Robust Stochastic Gradient Algorithm with an Adaptive Step-Size for Mobile Hands-Free Telephones," Proc. ICASSP'95, pp.1392-395, May 1995.

[14] T Gaňsler, M. Hansson, C. J. Ivarsson, and G. Salomonsson, "A Double-Talk Detector Based on Coherence," IEEE Trans. Commun. Vol.44, No.11, pp.1421-1427, Nov. 1996.

[15] J. H. Cho, D. R. Morgan, and J. Benesty, "An Objective Technique for Evaluating Doubletalk Detectors in Acoustic Echo Cancelers," IEEE Trans. SAP, Vol. 7, No. 6, pp.718-724, Nov. 1999.

[16] J. Benesty, D. R. Morgan, and J. H. Cho, "A New Class of Doubletalk Detectors Based on Cross-Correlation," IEEE Trans. SAP., vol.8, No.2, pp.168-172, Mar. 2000.

[17] S. J. Park, C. G. Cho, C. Lee, and D. H. Youn, "Integrated Echo and Noise Canceler for Hands-Free Applications," IEEE Trans. CAS-II, Vol.49, No.3, pp.188-195, Mar. 2002.

[18] R. Botros, O. Abdel-Alim, and P. Damaske, "Stereophonic Speech Teleconferencing," Proc. ICASSP'86, pp.1321-1324, Apr. 1986.

[19] T. Fujii and S. Shimada, "A note on Multi-Channel Echo Cancellers," Technical Report of IEICE, CS84-178, pp. 7-14, Jan. 1985 (in Japanese).

[20] A. Hirano and A. Sugiyama, "A New Multi-Channel Echo Canceller with a Single Adaptive Filter per Channel," Proc. Nat. Conv. of IEICEJ, A-202, Mar 1991.

[21] A. Hirano and S. Koike, "Convergence Analysis of a Stereophonic Acoustic Echo Canceller Part I: Convergence Characteristics of Tap Weights," Proc. of 11th DSP Symposium of IEICEJ, pp. 569-574, Nov. 1996.

[22] M. M. Sondhi, D. R. Morgan, and J. L. Hall, "Stereophonic Acoustic Echo Cancellation– An Overview of the Fundamental Problem," IEEE SP Let., Vol.2, No.8, pp.148-151, Aug. 1995.

[23] J. Benesty, F. Amand, A. Gilloire and Y. Grenier, ' 'Adaptive Filtering Algorithms for Stereophonic Echo Cancellation," Proc. of ICASSP'95, pp. 3027-3030, May 1995.

[24] S. Shimauchi and S. Makino, "Stereo Projection Echo Canceller with True Echo Path Estimation," Proc. of ICASSP'95, pp. 3059-3062, May 1995.

[25] T. Gänsler and P. Eneroth, "Influence of Audio Coding on Stereophonic Acoustic Echo Cancellation," Proc. of ICASSP'98, pp. 3649-3652, May 1998.

[26] A. Gilloire and V. Turbin, "Using Auditory Properties to Improve the Behavior of Stereophonic Acoustic Echo Cancellers," Proc. of ICASSP'98, pp. 3681-3683, May 1998.

[27] Y.-W. Jung, J.-H. Lee, Y.-C. Park, D.-H. Youn, "A New Adaptive Algorithm for Stereophonic Acoustic Echo Canceller," Proc. of ICASSP2000, pp. 801-804, Jun. 2000.

[28] J. Benesty, D. R. Morgan and M. M. Sondhi, "A Better Understanding and an Improved Solution to the Problems of Stereophonic Acoustic Echo Cancellation," IEEE Trans. SAP, Vol. 6, No. 2, pp.156-165, Mar. 1998.

[29] A. Hirano, K. Nakayama, D. Someda, and M. Tanaka, "Stereophonic Acoustic Echo Canceller without Pre-Processing," Proc. ICASSP2004, May 2004.

[30] J. Benesty, D. R. Morgan and M. M. Sondhi, "A Hybrid Mono/Stereo Acoustic Echo Canceler," IEEE Trans. SAP, Vol. 6, No. 5, pp.468-475, Sep. 1998.

[31] T. Gaňsler and J. Benesty, "New Insights into the Stereophonic Acoustic Echo Cancellation Problem and an Adaptive Nonlinearity Solution," IEEE Trans. SAP., Vol.10, No.5, pp.257-267, Jul. 2002.

[32] D. R. Morgan, J. L. Hall, and J. Benesty, " Investigation of Several Types of Nonlinearities for Use in Stereo Acoustic Echo Cancellation," IEEE Trans. SAP., Vol.9, No.6, pp.686-696, Sep. 2001.

[33] Y. Joncour, A. Sugiyama, and A. Hirano, "A Stereo Echo Canceler with Correct Echo-Path Identification and Good Sound Localization," Proc. DSP Workshop, Aug. 1998.

[34] A. Sugiyama, Y. Joncour, and A. Hirano, "A Stereo Echo Canceler with Correct Echo-Path Identification Based on an Input Sliding Technique,h IEEE Trans. SP, pp.2577-2587, Nov. 2001.

[35] M. Ali, "Stereophonic Acoustic Echo Cancellation System Using Time-Varying All-Pass Filtering for Signal Decorrelation," Proc. of ICASSP'98, pp. 3689-3692, May 1998.

[36] A. Hirano, K. Nakayama and K. Watanabe, "Convergence Analysis of Stereophonic Echo Canceller with Pre-Processing - Relation between Pre-Processing and Convergence -," Proc. ICASSP 1999, pp.861-864, Mar. 1999.

[37] A. Hirano, K. Nakayama, K. Takebe, "Stereophonic Acoustic Echo Canceller with Pre-Processing - Second-Order Pre-Processing Filter and Its Convergence -," Proc. IWAENC2003, pp.63-66, Sep. 2003.

[38] Y. Joncour and A. Sugiyama, "A New Stereo Echo Canceler with Pre-Processing Combined with Nonlinear Transformations," Proc. DSP Symp. of IEICEJ, pp.183-188, Nov. 1997.

[39] S. Bergman, C. Grewin and T. Rydén, "The SR Report on The MPEG/Audio Subjective Listening Test Stockholm April/May 1991," ISO/IEC JTC1/SC2/WG11 MPEG 91/010, May 1991.

[40] CCIR Recommendation 562, 1990.