AN ECHO CANCELLER USING SMOOTHED-COEFFICIENT FILTER WITH ADAPTIVE TIME CONSTANT CONTROLLED BY HIGH-PASS ERRORS

Osamu Hoshuyama

Common Platform Software Research Laboratories
NEC Corporation

ABSTRACT

This paper proposes a double-talk-robust echo canceller using a smoothed-coefficient filter (SCF). The SCF's tap coefficients are obtained by smoothing those of a pilot adaptive filter (PAF) for better accuracy and stability. The smoothing time constant is controlled by comparing mean squared high-pass errors of the SCF and the PAF. The high-pass filters enable precise control of the SCF by eliminating low frequency components of the near-end signal which cause comparison errors. Simulation results in a hands-free talk scenario demonstrate that the proposed echo canceller achieves robust echo cancellation even in severe double-talk periods.

Index Terms— Echo cancellation, Adaptive filters, Speech enhancement, Acoustic signal processing.

1. INTRODUCTION

In an acoustic echo canceller, double-talk control is an important function [1, 2]. If the control fails, the tap coefficients of the adaptive filter may be perturbed by the near-end signal, sometimes resulting in a howling or screaming noise.

For the double-talk robustness, echo cancellers using a pilot adaptive filter (PAF) and a main filter with copied tap coefficients have been proposed [3, 4, 5]. They detect double-talk periods based on the mean squared errors (MSEs) of the PAF and the copied-coefficient filter (CCF). When the MSE of the PAF is smaller than that of the CCF, which is considered as a single-talk state, the PAF's coefficients are copied to the CCF. In case of double talk, the MSE of the PAF is larger than that of the CCF. This means that the PAF's coefficients have large errors, therefore, the CCF holds its tap coefficients, resulting in double-talk robustness.

However, the performance of the echo cancellers using the CCF deeply depends on the double-talk detector. Just one wrong copy during a double-talk period causes fatal degradation of the echo cancellation performance. Therefore, the echo cancellers using the CCF can not track echo-path changes during double-talk periods.

The double-talk detection method itself has a problem that the comparison results of the MSEs are not so reliable as expected. Especially, when the near-end signal includes large low-frequency components, the high short-term autocorrelation in the near-end signal may make the magnitude relation of the MSEs reversed frequently even after the PAF has converged, resulting in false double-talk detection.

This paper proposes a new echo canceller structure using a smoothed coefficient filter (SCF) with adaptive time constant controlled by high-pass errors. The coefficient smoothing can continuously control the robustness against the false double-talk detection, and enables tracking to the echo-path changes during double-talk periods. It also increases the reliability of MSE comparison results. The high-pass errors have less low frequency components, leading to more accurate double-talk detection. The robust structure with the accurate double-talk detection achieves desirable echo cancellation in double-talk periods.

2. CONVENTIONAL ECHO CANCELLER WITH COPIED-COEFFICIENT FILTER

An echo canceller using a CCF shown in Fig. 1 is mathematically described as:

$$\begin{aligned} X(k) &\equiv [x(k), x(k-1), ..., x(k-N+1)]^T, && (1)\\ y_{\text{PAF}}(k) &= W_{\text{PAF}}^T(k)\, X(k), && (2)\\ y_{\text{CCF}}(k) &= W_{\text{CCF}}^T(k)\, X(k), && (3)\\ e_{\text{PAF}}(k) &= d(k) - y_{\text{PAF}}(k), && (4)\\ e_{\text{CCF}}(k) &= d(k) - y_{\text{CCF}}(k), && (5) \end{aligned}$$

where k is the time index, and $X(k)$ is the input signal vector consisting of the far-end signal samples $x(k)$. $d(k)$ is the microphone signal consisting of the echo and the near-end signal. $y_{\text{PAF}}(k)$, $y_{\text{CCF}}(k)$, $e_{\text{PAF}}(k)$, and $e_{\text{CCF}}(k)$ are the output signals and the output errors of the PAF and the CCF, respectively. $W_{\text{PAF}}(k)$ and $W_{\text{CCF}}(k)$ are the N-tap coefficient vectors in the PAF and the CCF, respectively.

When the normalized least-mean-squares algorithm is used, its tap-coefficient update is expressed as follows:

$$\begin{aligned} W_{\text{PAF}}(k+1) &= W_{\text{PAF}}(k)\\ &\quad +\alpha\, e_{\text{PAF}}(k)\, X(k)\, [X^T(k)X(k)+\delta]^{-1}, \qquad (6) \end{aligned}$$

where α is the global stepsize which dominates tracking speed, and the final error. A small positive number δ is for regularization.

Update of the CCF is performed as follows.

$$W_{\text{CCF}}(k+1) = \begin{cases} W_{\text{PAF}}(k) & \text{If Single Talk} \\ W_{\text{CCF}}(k) & \text{If Double Talk} \end{cases}. \qquad (7)$$

The detection of single talk or double talk is based on comparison of the MSEs.

$$\begin{aligned} \text{MSE}_{\text{PAF}}(k) &= \gamma\, e_{\text{PAF}}^2(k) + (1-\gamma)\,\text{MSE}_{\text{PAF}}(k-1), && (8)\\ \text{MSE}_{\text{CCF}}(k) &= \gamma\, e_{\text{CCF}}^2(k) + (1-\gamma)\,\text{MSE}_{\text{CCF}}(k-1), && (9) \end{aligned}$$

where γ is a coefficient to determine the time constant for averaging squared errors.

In Ochiai's structure, double-talk periods are detected based on the theory that the expectations of MSE_{PAF} and MSE_{CCF} change their magnitude relation when MSE_{PAF} converged to the floor determined by the near-end signal. In case of a single talk, MSE_{PAF} still does not reach the floor, therefore, it is smaller than MSE_{CCF}. On the other hand, during a double-talk period, MSE_{CCF} is smaller than MSE_{PAF}. When a single-talk period is detected, W_{PAF} is copied to W_{CCF}, and when with a detection of double talk, the CCF holds its

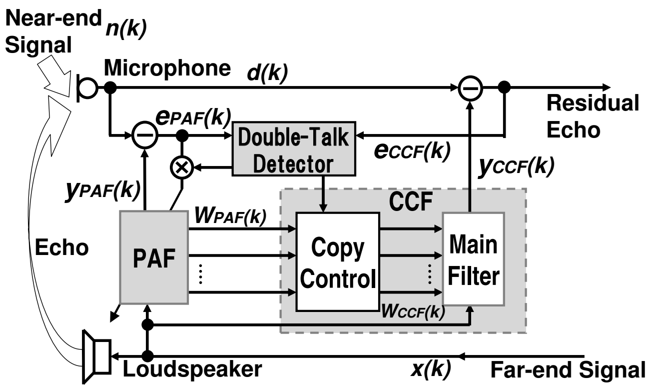

Fig. 1. Echo Canceller Using CCF.

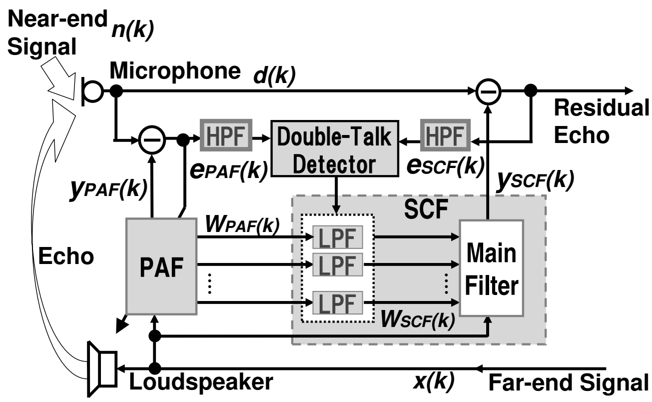

Fig. 2. Proposed Echo Canceller.

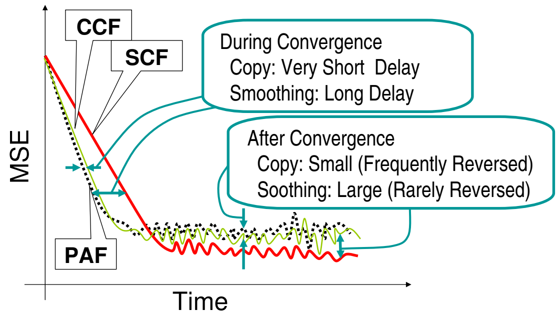

Fig. 3. Typical Convergence Curves of MSEs for PAF, CCF, and SCF.

coefficients. This structure, which was proposed for line echo cancellation, can not be directly applied for acoustic echo cancellation, because it is very sensitive to false single-talk detections. Just one wrong copy during a double-talk period causes fatal degradation of the echo cancellation performance. When an echo-path change is detected during a double-talk period, W_{PAF} are already perturbed by the near-end signal. If W_{PAF} are copied to the CCF, W_{CCF} will have large errors, resulting in awful echo or a howling in the worst case. Therefore, tracking to the echo-path changes during double-talk periods is almost impossible.

Haneda et al. proposed an improved structure with a sophisticated copy control algorithm for acoustic echo cancellation, although its details are not disclosed [4]. The copy control algorithm is very conservative, because even when detection of echo-path change is delayed, W_{CCF} can immediately catch up with W_{PAF} [6]. Haneda's structure has shown good performance in most conditions, however, the conservative copy control could cause slow tracking.

There is another problem in the double-talk detection method based on MSE comparison. Just the comparison of the MSEs is not reliable, and false double-talk detections occur frequently. Especially, when the near-end signal is loud speech or low-frequency noise, it has a high short-term autocorrelation due to the low frequency components or the periodical components. The high autocorrelation causes a bias which makes $\mathrm{MSE}_{\mathrm{PAF}}$ smaller [7]. Because the difference of the MSEs is very small, magnitude relation of the MSEs may be reversed, and a false double-talk detection may occur. Even an improved detection method [5] is directly influenced by the high short-term autocorrelation.

3. PROPOSED ECHO CANCELLER

The proposed echo canceller shown in Fig. 2 uses a SCF with adaptive time constant. The SCF is a main filter with tap coefficients smoothed by low-pass filters (LPFs) from the PAF's coefficients. The time constant for smoothing in the LPFs is controlled based on the comparison results between the MSEs of the SCF and the PAF. The MSEs are calculated from high-pass errors.

The LPFs for coefficient smoothing are realized by infinite impulse response (IIR) filters, and the MSE of the SCF is calculated as follows:

$$W_{\mathrm{SCF}}(k+1) = \eta(k)\, W_{\mathrm{PAF}}(k) + \{1-\eta(k)\}\, W_{\mathrm{SCF}}(k), \quad (10)$$
$$= \eta(k)\, \{W_{\mathrm{PAF}}(k) - W_{\mathrm{SCF}}(k)\} + W_{\mathrm{SCF}}(k), \quad (11)$$
$$\mathrm{MSE}_{\mathrm{SCF}}(k) = \gamma\, e^2_{\mathrm{SCF}}(k) + (1-\gamma)\, \mathrm{MSE}_{\mathrm{SCF}}(k-1), \quad (12)$$

where $\eta(k)$ is a forgetting factor to determine the time constant of coefficient smoothing. The time constant in the SCF is controlled in a simple way based on the MSE comparison between the PAF and the SCF as follows:

$$\eta(k) = \begin{cases} \eta_1 & \text{If } \mathrm{MSE}_{\mathrm{SCF}}(k) > \mathrm{MSE}_{\mathrm{PAF}}(k) \\ & \quad : \text{Single Talk} \\ \eta_2 & \text{If } \mathrm{MSE}_{\mathrm{SCF}}(k) \le \mathrm{MSE}_{\mathrm{PAF}}(k) \\ & \quad : \text{Double Talk} \end{cases} \quad (13)$$

In this paper only the time constant is controlled to show the preferable behavior of the SCF, however, for a better convergence behavior, the stepsize α can also be controlled.

For the high-pass filters (HPFs) to calculate the MSEs, any high-pass/low-cut functions do the job. In this paper, the following transfer function is used:

$$HPF(z) = 1 - 2\,z^{-1} + z^{-2}. \quad (14)$$

3.1. Features of SCF

Figure 3 shows typical convergence curves of the MSEs for a PAF, a CCF, and an SCF. The smoothing in the SCF leads to less variance, which means more correct tap coefficients as far as the tap coefficients identified by the PAF is unbiased [8]. [1]

When the SCF is used with a double-talk controller based on the MSE comparison, the advantages over the CCF are as follows:

[1] This is reasonable because the adaptation algorithms can be interpreted as IIR LPFs with time varying smoothing time constant comprising of the input signal components [9]. In the proposed echo canceller, the additional IIR LPFs are located outside of the adaptation loop, therefore, it has no stability problem like [10, 11].

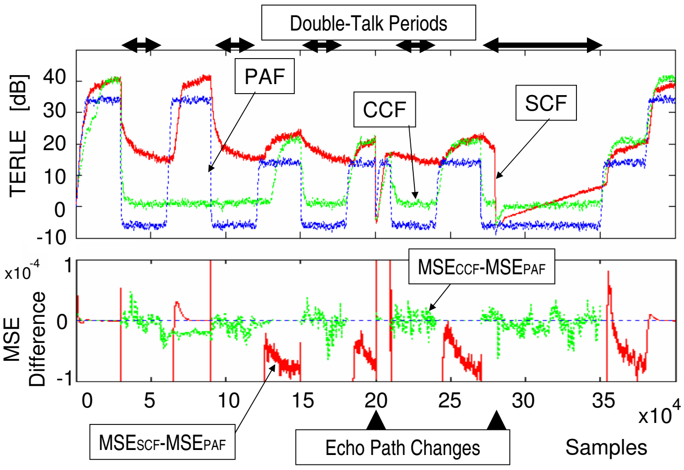

Fig. 4. TERLEs and MSE Differences in Environment with Double Talk and Echo-Path Changes.

1. By appropriate setting of η_1 and η_2, the influence on the SCF by a false double-talk control can be much smaller than that on the CCF. Actually, the CCF can be considered as a special case of SCF with $\eta_1 = 1.0$ and $\eta_2 = 0.0$.

2. By appropriate setting of η_2, the SCF can track the echo-path changes even in double-talk periods.

3. During PAF's convergence, group delay caused by the smoothing IIR filter can be much longer than that with copying. The long group delay increases the MSE difference. (Fig. 3)

4. After the PAF's convergence, the MSE difference is larger because the SCF's tap coefficients have significantly less variance. (Fig. 3)

5. The large MSE difference leads to stable comparison results and reliable double-talk detection. Therefore, a simple control method achieves sufficiently good echo cancellation.

The increase of computations with the SCF is affordable for the latest signal processors. When W_{SCF} is updated by (11), the additional computations by an SCF compared to a CCF are approximately N multiplications and $2N$ additions.

3.2. Effectiveness of High-Pass Filters

High-pass filters to obtain MSEs remove the low-frequency components which dominate the short-term autocorrelation of the near-end signal. Although the MSEs of the high-pass errors are smaller than those without the high-pass filters, the bias in the PAF's MSE caused by the short-term autocorrelation of the near-end signal is more reduced. Therefore, magnitude relation (comparison result) becomes more stable, and the false double-talk detections are reduced.

4. SIMULATIONS

Simulations to evaluate the proposed echo canceller have been performed assuming acoustic echo cancellation for hands-free phones. Artificial environments were used for visibility of the simulation results.

4.1. Double-Talk Robustness

To demonstrate the echo cancellation performance in double-talk periods, the true echo return loss enhancement (TERLE) defined by the following equation was measured for the CCF in the conventional echo canceller (hereinafter "CCF"), the PAF and the SCF in the proposed echo canceller (hereinafter "PAF" and "SCF").

$$\mathrm{TERLE}(k) = -10\log_{10}\left\{\frac{\overline{[e(k)-n(k)]^2}}{\overline{[d(k)-n(k)]^2}}\right\}, \qquad (15)$$

where $n(k)$ is the near-end signal, and $\overline{[\cdot]}$ is averaging over 200 samples.

The signals in an environment with severe double talk and echo-path changes were generated from white Gaussian noises. The far-end signal had a constant amplitude, and the echo paths were selected so that the echo level was also constant. The near-end signal was a burst noise with variable levels, which imitates the bursty characteristics of the near-end speech. The number of tap coefficients, N, was 270, and the sampling rate was 8 kHz. Echo-to-near-end-signal ratio was about 0dB during the severe double talk periods. Other parameters were: $\alpha = 0.4$, $\eta_1 = 0.0005$, $\eta_2 = 0.00002$, $\gamma = 0.001$, and $\delta = 0.00001$. The double-talk detector for the copy controller in the CCF expressed by (7) used the same criterion as in (13) because the details of the controller is not disclosed in [4].

$$\begin{aligned} &\text{Single Talk} \;:\; \text{If } \mathrm{MSE}_{\mathrm{CCF}} > \mathrm{MSE}_{\mathrm{PAF}}, \\ &\text{Double Talk} \;:\; \text{If } \mathrm{MSE}_{\mathrm{CCF}} \le \mathrm{MSE}_{\mathrm{PAF}}. \end{aligned} \qquad (16)$$

This is not fair in terms of double-talk robustness comparison, however, the author thinks that the simulations are fair to show the difficulty of the copy controller and superiority of the SCF over the CCF.

TERLEs for the PAF, CCF, and SCF, and MSE differences are shown in Fig. 4. For the TERLEs, a higher value is desired. As for the MSE differences, a single-talk period is detected when a MSE difference is higher than 0. Ideally, a single-talk period should be detected only when the echo-path changes or when the noise level sufficiently decreases.

At first, there is almost no difference. However, the CCF shows slow convergence at 1×10^4th samples due to a false double-talk detection. In the 1st double-talk period from 3×10^4th sample to 6×10^4th sample, the near-end signal increased to imitate a double talk. The PAF was perturbed by the large near-end signal, and its TERLE decreases. The CCF failed at detecting the double talk, resulting in decrease of TERLE. On the other hand, the SCF succeeded at double-talk detection, and appropriately reduced the smoothing time constant, therefore, obtained a TERLE of 15 dB, which is 20 dB better than that by the PAF.

After the 1st double-talk period, from 6×10^4th sample to 9×10^4th sample, the near-end signal decreased by 40 dB. The PAF immediately updated its tap coefficients. The CCF again failed at detecting the convergence condition, and could not update the CCF tap coefficients. The SCF had a delay at detecting the convergence condition, however, at the moment when the MSE of the PAF exceeded that of the SCF, the SCF began to follow the PAF by setting $\eta(k)$ to a large value η_1. After the PAF converged, $\eta(k)$ was again set to a small value η_2 for accurate tap coefficients.

From 9×10^4th sample to 20×10^4th sample, the near-end signal level varied. In all the conditions, the SCF keeps the highest TERLE. At 20×10^4th sample, echo path changed with small near-end signal. All the methods detecded the change, and showed fast tracking to the echo-path change. The SCF was the slowest, however, the delay was not significantly audible.

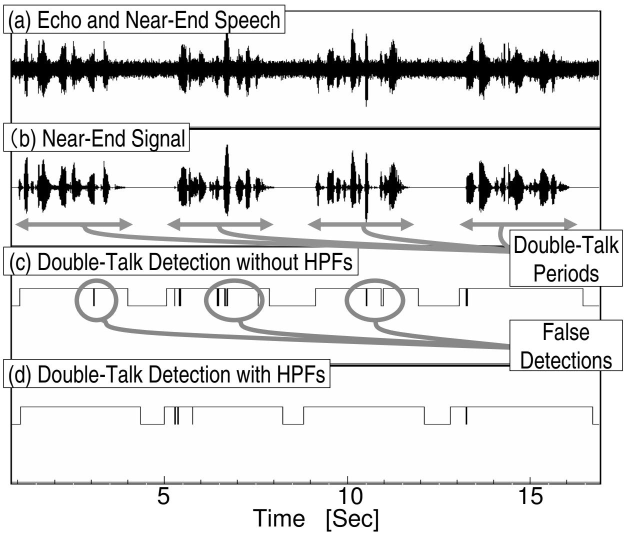

Fig. 5. Double-Talk Detection with and without HPFs.

In the 5th double-talk period from $27{\times}10^4$th sample to $35{\times}10^4$th sample, at $28{\times}10^4$th sample, the echo path changed again. The PAF converged to a low TERLE due to the double talk. The CCF could not update its tap coefficients. However, the SCF tracked the echo-path change. The convergence curve is comparable to that with a PAF with a stepsize of 0.02. It is not so fast, however, is a best effort for the simple control method.

4.2. Double-Talk Detection Performance

Simulations to demonstrate the effectiveness of the HPFs for MSE calculation were also performed. Figure 5 shows the signal waveforms and double-talk detection results of the proposed echo canceller with and without the HPFs. The far-end signal was white Gaussian noise, and the near-end signal was a sequence of speech. Other parameters were the same as those in Sec. 4.1.

In Fig. 5, (a) shows the waveform of the microphone signal which consists of the echo and the near-end speech, and (b) shows only the near-end signal waveform. Ideally, the periods indicated by arrows in (b) should be detected as double-talk periods. Figures 5(c) and (d) show the detection results without and with the HPFs, respectively. A high level indicates detected double talk, and a low level means single talk. In the areas of ovals in (c), the double-talk periods are misdetected as single talk periods. These false detections occur when the near-end signal has large low frequency components or has clear harmonic structure, which means a high autocorrelation. On the other hand in (d), some of the false detections in (c) disappeared, which is achieved by removing the low frequency components in the HPFs.

It should be noted that this good double-talk detection performance in the proposed echo canceller is achieved by just a simple comparison of the MSEs. The proposed echo canceller can provide stable and reliable echo replica, which contributed to the sound quality of the nonlinear echo canceller in [12]. In this paper, only simple simulation results have been presented for visibility of the figures, however, the proposed echo canceller works properly in the real world with tough noises and distortion.

5. CONCLUSIONS

This paper has proposed a new double-talk-robust echo canceller using a SCF with an adaptive time constant controlled by high-pass errors. The SCF's tap coefficients are obtained by smoothing those of a PAF for better accuracy and stability. The smoothing time constant is controlled by comparing mean squared high-pass errors of the SCF and the PAF. The high-pass filters enable precise control of the SCF by eliminating low frequency components of the near-end signal which cause false double-talk detection. Simulation results in a hands-free talk scenario have demonstrated that the proposed echo canceller achieves robust echo cancellation even in a severe double-talk condition.

ACKNOWLEDGEMENT

The author would like to thank Dr. Akihiko Sugiyama, Mr. Toshiyuki Nomura, and Dr. Masahiro Serizawa of Media Processing Technology Group, Common Platform Software Research Laboratories, NEC Corporation for their encouragement and valuable comments on this research.

6. REFERENCES

[1] E. Hänsler, and G. Schmidt eds., *Acoustic Echo and Noise Control: A Practical Approach*, Wiley, 2004.

[2] J.-M. Valin, and I. Collings, "A New Robust Frequency Domain Echo Canceller with Closed-Loop Learning Rate Adaptation," IEEE Proc. ICASSP2007, pp. I-93–96, May 2007.

[3] K. Ochiai, T. Araseki, and T. Ogihara, "Echo Canceller with Two Echo Path Models," IEEE Trans. Comn, Vol. COM-25, No. 6, pp. 589–595, Jun. 1977.

[4] Y. Haneda, S. Makino, J. Kojima, and S. Shimauchi, "Implementation and Evaluation of an Acoustic Echo Canceller using Duo-Filter Control System," Proc. EUSIPCO96, pp. 1115–1118, Sep. 1996.

[5] M. A. Iqbal, and S. L. Grant, "Novel and Efficient Download Test for Two Path Echo Canceller," IEEE Proc. WASPAA2007, pp. 167–170, Oct, 2007.

[6] Y. Haneda, Personal Communications, 2007.

[7] A. A. Beex, and J. R. Zeidler, "Steady-State Dynamic Weight Behavior in (N)LMS Adaptive Filters," Chapter 9 in *Least-Mean-Square Adaptive Filters*, pp. 335–443, Willey, 2003.

[8] M. Niedzwiecki, "Identification of Time-Varying Systems Using Combined Parameter Estimation and Filtering," IEEE Trans. ASSP, Vol. 38, No. 4, Apr. 1990.

[9] K. Fujii, Y. Sakai, and J. Ohga, "Influence of Time Constant on Coefficients Estimation Error Derived from a Lowpass Filter Expression for Learning Identification Algorithm," IEICE Elec. and Comn. in Japan (Pt. 3: Fund. Elec. Sci.) Vol. 74, No. 12, pp. 43–54, Willey, Dec. 1991.

[10] S. Roy, and J. J. Shynk, "Analysis of the Momentum LMS Algorithm," IEEE Trans. ASSP, Vol.38, No. 12, pp. 2088–2098, Dec. 1990.

[11] S. Kinjo, and H. Ochi, "A New Robust Adaptive Filter for Colored Signal Input," IEICE Trans. Fundamentals, Vol. E-78-A, No.3, Mar. 1995.

[12] O. Hoshuyama, and A. Sugiyama, "An Acoustic Echo Suppressor Based on a Frequency-Domain Model of Highly Nonlinear Residual Echo," IEEE Proc. ICASSP2006, pp. V-269–272, May 2006.