Acoustic Echo Canceller

White Paper

June 2011

1. Introduction

Acoustic Echo

Communication technology has advanced to the point where teleconferences that use communication lines to connect remote conference rooms are taking place every day.

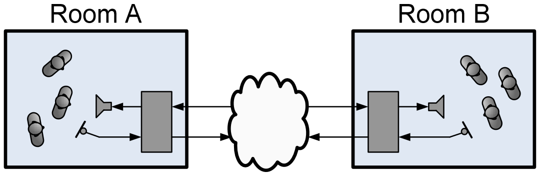

Fig.1 Teleconference

In a teleconference, as in Fig.1, sound is exchanged in both directions through communication lines. When the voice of the person speaking in room A is transferred to room B through communication lines and reproduced by a loudspeaker, that voice can be heard in room B. However the microphone in room B also picks up the voice of the person speaking in room A and sends that voice back to room A. Because of this, in room A, the voice of the person speaking is reproduced by the loudspeaker as an echo. This is referred to as acoustic echo. In addition to hindering conversation, it can cause feedback. To have a smooth conference, it is necessary to reliably suppress acoustic echo.

Acoustic Echo Canceller

One way to suppress acoustic echo is, for example, to reduce the signal level of the microphone in room B when a person is speaking in room A. This is referred to as echo suppression. However, when echo suppression is used and people in room A and room B talk at the same time (double talk), their voices get cut off. This makes it difficult for people to hear what was said and prevents smooth conversation. To solve this problem, in 1966, M. M. Sondhi [1][2] proposed an acoustic echo canceller with an ADF (adaptive digital filter). However, to actually create an acoustic echo canceller with an ADF, one would need highly advanced signal processing technology and a signal processing device with high computational capabilities. In recent years, the advancement of signal processing technology and DSPs (digital signal processors) has made it possible to create sophisticated echo cancellers.

2. Adaptive Digital Filter

An ADF is a digital filter that can learn. ADFs analyze input and output signals in real time and identify unknown transfer systems [3].

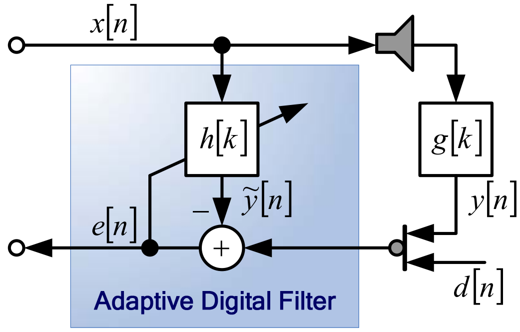

Fig.2 Acoustic Echo Canceller Using ADF

An acoustic echo canceller that uses an ADF is configured as in Fig.2. The ADF is used to simulate the transfer system from the loudspeaker to the microphone, $g[k]$, using a filter, $h[k]$. To remove acoustic echo $y[n]$, the echo canceller obtains pseudo echo signal $\tilde{y}[n]$ by sending the input signal from the far end, $x[n]$, to filter $h[k]$ and subtracting the pseudo echo signal from the microphone signal. The signal with the acoustic echo removed from it, $e[n]$, is not just sent to the far end; the ADF also learns from it. The ADF learns how to minimize the errors, $e[n]$, in its filtering. The algorithms that are used for this learning are referred to as adaptive algorithms. There are many different types of these algorithms. Factors such as the estimation accuracy, convergence rate, and amount of computation vary depending on the algorithm that is used. One of the best known adaptive algorithms is the NLMS (normalized least mean square) algorithm. It updates the filter in accordance with the following formula. [3]

$$h_{n+1}[k] = h_n[k] - \frac{\mu}{\sigma_x^2} \cdot x[n-k] \cdot e[n]$$

3. Implementation

In an actual teleconference, the microphone signal doesn't just include acoustic echo; it also includes external noise $d[n]$, which includes background noise and the sound of the near-end speaker. External noise interferes with the learning of the filter and reduces the filter's estimation accuracy. Also, time variation and nonlinearity in the transfer system, $g[k]$, also reduce the estimation accuracy and result in residual echo. To supplement the functionality of the ADF, an echo suppressor for suppressing residual echo is connected in the later stage of the ADF. (**Fig.3**)

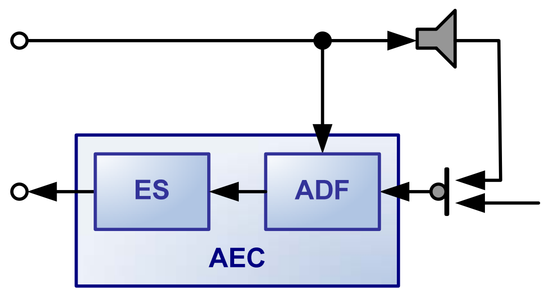

Fig.3 Acoustic Echo Canceller

The acoustic echo canceller shown in **Fig.3** is implemented in a Yamaha DSP that is specially designed for audio signal processing, the YSS950 DAP1 (32-bit floating-point; shown in Fig.4). With a sampling rate of 48 kHz, this DSP makes it possible to obtain acoustic echo cancelling with both high quality sound and high precision.

Fig.4 Yamaha YSS950 DAP1

4. Specification

4.1. Measurement

We evaluated the acoustic echo canceller implemented in chapter 3 in a conference room with a maximum capacity of 180 people and a reverberation time of approximately 1.0 s (Fig.5). In the teleconferencing system, we used front loudspeakers (the Yamaha IF2108), ceiling loudspeakers, and a microphone (the Shure SM93). From the loudspeakers, we output a signal combining the signal received from the far end and the microphone signal. We performed AEC processing on the microphone signal and sent it to the far end. The amount of acoustic echo is expressed as the difference between the level of the input signal from the far end and the level of the acoustic echo and is referred to as ERL (echo return loss). The larger the ERL, the smaller the acoustic echo and the more advantageous the installation conditions. It is best to install a teleconferencing system so that ERL is greater than 10 dB. We conducted this evaluation with the inherent ERL at -1.2 dB, 5.5 dB, and 11.5 dB. An ERL value of -1.2 dB would constitute an extremely strong acoustic echo.

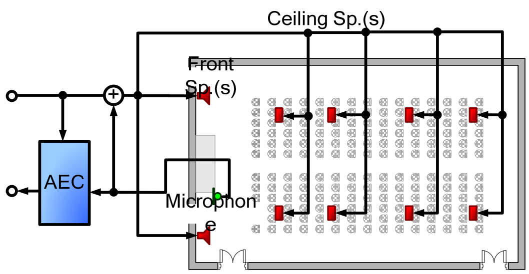

Fig.5 Setup

To evaluate echo suppression for single talk (in which only the person speaking at the far end is talking), we used a CS signal [4], and measured TCLw (weighted Terminal Coupling Loss) [5]. To evaluate conversation performance for double talk, we used a CS signal and an AMFM signal [4] and measured the amount of echo suppression (TCL_{wdt}) and the amount of insertion loss ($A_{H,S,DT}$) for the sound of the near-end speaker [6].

4.2. Result

Performance Evaluation for Single Talk

Fig.6 is the amount of echo suppression during single talk (TCLw). ITU-T Rec. 341 [7] recommends that echo suppression be not less than 35 dB, but regardless of the amount of ERL, a high amount of echo suppression, approximately 70 dB, is achieved. It is clear that the acoustic echo canceller can reliably remove acoustic echo.

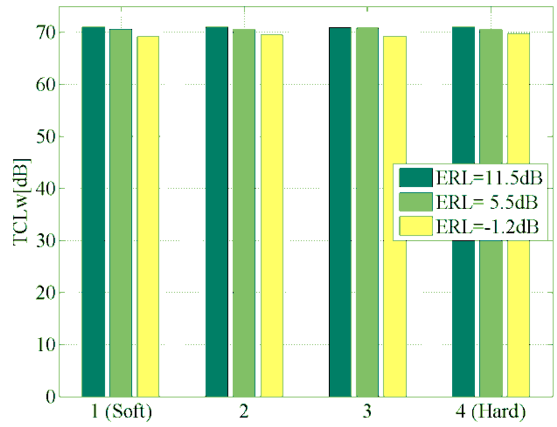

Fig.6 TCLw

Performance Evaluation for Double Talk

Fig.7 is the amount of echo suppression during double talk (TCL_{wdt}). During double talk, because both speakers are talking at the same time, acoustic echo stands out less than it does with single talk. The necessary amount of echo suppression varies depending on the volume of the sound being produced by the loudspeakers, the sensitivity of the microphone, and other settings, but given the measurement conditions of this evaluation, it is presumed that echo suppression of not less than 15 to 25 dB is necessary (ITU-T Rec. P.340[8]). When the amount of acoustic echo is low (ERL = 11.5 dB), sufficient echo suppression is achieved even when AEC is set to type 1 (softest). When the amount of acoustic echo is high (ERL = -1.2 dB), it is necessary to set AEC to 2 or higher to suppress the acoustic echo.

Fig.8 shows the amount of insertion loss ($A_{H,S,DT}$) for the sound of the near-end speaker during double talk. When the insertion loss is great, the possibility that the sound will be cut off or that the quality of the sound will change increases. According to ITU-T Rec. P.340 [8], an insertion loss of not more than 3 dB is recommended. Even when the amount of acoustic echo is high (ERL=

-1.2 dB), if AEC is set to a value not greater than 2, the influence to the near-end speaker is reduced, and conversations can be conducted smoothly.

The results of the performance evaluation for double talk show us that when ERL is small (the amount of acoustic echo is high), there is a tradeoff between the amount of echo suppression and the sound quality. Therefore, to ensure smooth conversation, it is better to use a configuration in which ERL is sufficiently high (acoustic echo is low).

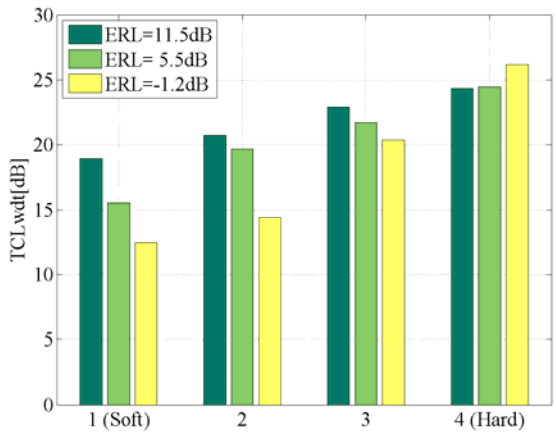

Fig.7 TCL_{wdt}

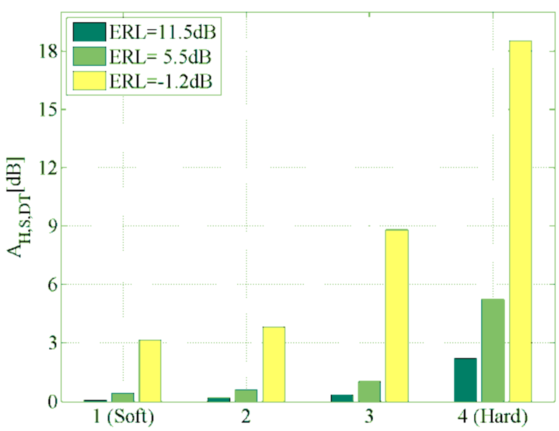

Fig.8 $A_{H,S,DT}$

5. Conclusion

The results of the evaluation show that an acoustic echo canceller can reliably suppress echoing and enable smooth conversation. However, during double talk, when the ERL is low (the amount of acoustic echo is high), it is more likely that problems will occur, such as the sound being cut off and the sound quality changing. Therefore, one should appropriately install microphones and loudspeakers so as to reduce the generation of acoustic echo and then appropriately configure the acoustic echo canceller. By doing so, one can ensure a smooth teleconference.

6. Reference

[1] M. M. Sondhi, “Theory and Computer Simulation of a Self-Adapting Echo Canceller”, J. Acoust. Soc. Am, vol.40, Issue 5, p.1255, 1966.

[2] M. M. Sondhi, “An adaptive echo canceller”, Bell System Technical Journal, vol.XLVI, no.3, pp.497–510, March 1967.

[3] S. Haykin, “Adaptive Filter Theory”, 3rd Edition, Prentice-Hall, 1996.

[4] ITU-T Recommendation P.501 (06/07), “Test signals for use in telephonometry”

[5] ITU-T Recommendation G.122 (03/93), “Influence of national systems on stability and talker echo in international connections”

[6] ITU-T Recommendation P.1100 (10/2008), “Narrow-band hands-free communication in motor vehicles”

[7] ITU-T Recommendation P.341 (06/2005), “Transmission characteristics for wideband (150-7000 Hz) digital hands-free telephony terminals”

[8] ITU-T Recommendation P.340 (05/2000), “Transmission characteristics and speech quality parameters of hands-free terminals”

Sound & IT Development Division Hiraku Okumura
Sound Network Division Ryo Tanaka
Pro Audio Division Masanobu Ando

Yamaha Corporation

10-1 Nakasawa-cho, Naka-ku, Hamamatsu City,
Shizuoka Prefecture, Japan
http://proaudio.yamaha.co.jp