A HANDS-FREE UNIT WITH ADAPTIVE MICROPHONE ARRAY FOR DIRECTIONAL AGC

Kazunori Kobayashi, Yoichi Haneda, Ken'ichi Furuya, and Akitoshi Kataoka

kobayashi.kazunori@lab.ntt.co.jp
NTT Cyber Space Laboratories, NTT Corporation,
3-9-11 Midoricho, Musashino, Tokyo, 180-8585 JAPAN

ABSTRACT

We previously proposed a directional automatic gain controller (AGC) using a microphone array to adjust the speech levels for multiple talkers individually. In this paper, we present an implementation it in a hands-free unit for teleconferencing systems and performance evaluation results. The unit detects talker directions and estimates their speech levels. It then controls the directivity pattern to achieve constant speech levels for all talkers. Even when several talkers are speaking at different distances from the unit, it can adjust their different speech levels to appropriate levels simultaneously. Experiments show that the unit provides effective gain control for multiple talkers.

1. INTRODUCTION

In teleconferencing systems, it is important to adjust speech signals to appropriate levels. However, the speech level picked up by a microphone varies with distance to the talker or utterance volume. Therefore, an automatic gain controller (AGC) is used to adjust the speech level.

A conventional AGC [2] [3] estimates the speech level and then adjusts it by applying the gain to the speech signal. However, in a teleconference with multiple participants, the speech levels picked up by the microphone changes frequently since the participants are different distances from the microphone and they have different utterance volumes. In such a case, the conventional AGC might not adjust the speech level effectively because the level estimation cannot track a level that changes frequently. Furthermore, there is a problem that the noise level increases with gain adjustment. In particular, when the talker is far from the microphone, the large noise that increases with the gain deteriorates the quality of the transmitted signal.

We solve these problems by a directional AGC using a microphone array [1]. The microphone array forms a directivity pattern to achieve a constant level for multiple talkers and suppresses stationary noise. The directivity pattern is automatically controlled by detecting the directions of talkers and estimating their speech levels. Furthermore, to suppress acoustic echoes in hands-free teleconferencing systems, we investigated the combination of the directional AGC and an acoustic echo canceller (AEC).

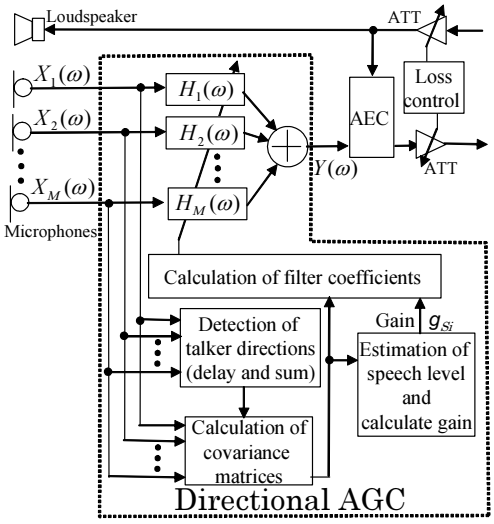

Figure 1: *Block diagram of the proposed hands-free system with directional AGC.*

In this paper, we implemented the directional AGC and AEC in a prototype teleconferencing unit including a loudspeaker and four microphones and evaluated its performance.

2. DIRECTIONAL AGC

A block diagram of the directional AGC is shown in Fig. 1. To estimate the speech level of each talker, the directional AGC detects the directions of talkers by using a delay-and-sum beamformer. The detector scans the beam toward all directions and a direction of maximum power is recognized as a talker. In this method, we assume that only one talker speaks at a given time. The directional AGC then stores the detected directions and their speech levels. From this information, the directivity pattern to achieve appropriate levels for all talkers is formed by using an adaptive beamforming method [4], as follows. The directional AGC applies filters to the signals received by

M microphones and then sums the results. The output signal $Y(\omega)$ is given by

$$Y(\omega) = \mathbf{X}(\omega)^{\mathrm{T}}\mathbf{H}(\omega), \quad (1)$$

where $\mathbf{X}(\omega)$ and $\mathbf{H}(\omega)$ are vectors of input signals received by the microphones and the filters, respectively. They are given by

$$\mathbf{X}(\omega) = (X_1(\omega), ..., X_M(\omega))^{\mathrm{T}}, \quad (2)$$

$$\mathbf{H}(\omega) = (H_1(\omega), ..., H_M(\omega))^{\mathrm{T}}. \quad (3)$$

While the i-th talker is speaking, the input signal $\mathbf{X}_{Si}(\omega)$ consists of speech signal $\mathbf{S}_i(\omega)$ and noise $\mathbf{N}(\omega)$. This is given by

$$\mathbf{X}_{Si}(\omega) = \mathbf{S}_i(\omega) + \mathbf{N}(\omega). \quad (4)$$

To adjust the speech levels of N talkers, the filters must ensure that the appropriate gains g_{Si} are applied to the speech components of all talkers. The conditions are given by

$$\mathbf{S}_i(\omega)^{\mathrm{T}}\mathbf{H}(\omega) = g_{Si} \cdot \mathbf{S}_i(\omega)^{\mathrm{T}}\mathbf{A}_{Si}, \quad \text{for } i = 1, ..., N \quad (5)$$

where $\mathbf{A}_{Si}$ is the mixing weight vector for the i-th talker:

$$\mathbf{A}_{Si} = (a_{Si,1}, ..., a_{Si,M})^{\mathrm{T}}. \quad (6)$$

The mixing weight vectors are preset based on the directions of microphones and talkers. For example, a large weight is set for the microphone whose directivity is suitable for the talker.

The condition for noise suppression is given by

$$\mathbf{N}(\omega)^{\mathrm{T}}\mathbf{H}(\omega) = 0. \quad (7)$$

We can obtain the filter coefficients by solving the simultaneous equations (5) and (7). The solution by the minimum squared error method is given by

$$\mathbf{H}(\omega) = \left(\sum_{i=1}^{N}\mathbf{R}_{SiSi}(\omega) + \mathbf{R}_{NN}(\omega)\right)^{-1} \cdot \sum_{i=1}^{N} g_{Si} \cdot \mathbf{R}_{SiSi}(\omega)\mathbf{A}_{Si}, \quad (8)$$

where $\mathbf{R}_{SiSi}(\omega)$ and $\mathbf{R}_{NN}(\omega)$ are covariance matrices of the i-th talker and the noise.

By detecting the noise period, we can obtain $\mathbf{R}_{NN}(\omega)$ using

$$\mathbf{R}_{NN}(\omega) = \overline{\mathbf{N}(\omega)\mathbf{N}(\omega)^{\mathrm{H}}}. \quad (9)$$

$\mathbf{R}_{SiSi}(\omega)$ is obtained while the i-th talker is speaking. The period is detected using estimates of talker directions. The noise component included in the covariance matrix can be cancelled by subtracting the noise covariance matrix. This is given by

$$\mathbf{R}_{SiSi}(\omega) = \overline{\mathbf{X}_{Si}(\omega)\mathbf{X}_{Si}(\omega)^{\mathrm{H}}} - \mathbf{R}_{NN}(\omega). \quad (10)$$

The gain g_{Si} is also obtained using covariance matrix $\mathbf{R}_{SiSi}(\omega)$. The speech level for the i-th talker P_{Si} is derived by

$$\begin{aligned} P_{Si} &= \sqrt{\int \overline{\left|\mathbf{S}_i(\omega)^{\mathrm{T}}\mathbf{A}_{Si}\right|^2} d\omega} \\ &= \sqrt{\int \mathbf{A}_{Si}^{\mathrm{T}}\mathbf{R}_{SiSi}(\omega)\mathbf{A}_{Si} d\omega}. \end{aligned} \quad (11)$$

The gain g_{Si} is then obtained by

$$g_{Si} = P_{opt}/P_{Si}. \quad (12)$$

The filter coefficients calculated by eq. (8) allow us to form a directivity pattern for directional AGC and noise suppression.

3. COMBINATION OF DIRECTIONAL AGC AND AEC

In a hands-free unit, an acoustic echo canceller (AEC) is necessary for full-duplex communication. In the combination of directional AGC and AEC, there is a problem that a change in echo path produced by the directional AGC negatively affects the performance of the AEC. If the AECs are used before the directional AGC, we can solve this problem. However, this solution is unsuitable for implementing on a low-cost DSP, because we need as many AECs as microphones. To reduce the adverse effect of the echo path change, we add the echo suppression condition to the filter design of the directional AGC. It is given by

$$\mathbf{E}(\omega)^{\mathrm{T}}\mathbf{H}(\omega) = 0, \quad (13)$$

where $\mathbf{E}(\omega)$ is the vector of echo signals received by the microphones.

We can obtain the filter coefficients by solving the simultaneous equations (5), (7), and (13);

$$\mathbf{H}(\omega) = \left(\sum_{i=1}^{N}\mathbf{R}_{SiSi}(\omega) + \mathbf{R}_{NN}(\omega) + \mathbf{R}_{EE}(\omega)\right)^{-1} \cdot \sum_{i=1}^{N} g_{Si} \cdot \mathbf{R}_{SiSi}(\omega)\mathbf{A}_{Si}, \quad (14)$$

where $\mathbf{R}_{EE}(\omega)$ is a covariance matrix of the echo signals.

By using the filter coefficients calculated by eq. (14), the echo can be reduced before the AEC. Thus, the problem of the echo path change can be reduced.

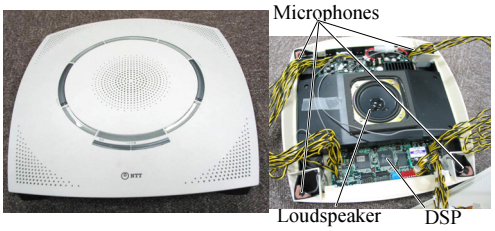

Figure 2: *External and internal views of the hands-free unit.*

4. IMPLEMENTATION

We implemented the directional AGC and AEC in a hands-free system with a fixed point DSP (TMS320C5416 from Texas Instruments). The DSP has calculation performance of 160 MIPS and 256 kbyte of on-chip RAM. A block diagram of the hands-free system is shown in Fig. 1. In this system, the sampling frequency of A/D or D/A converters is 16 kHz and four microphones are used for the directional AGC. The filter lengths of the directional AGC and AEC are 128 and 2048 taps, respectively. If there are many participants, the directional AGC controls the directivity pattern for only the three talkers that spoke most recently. This is because a covariance matrix that uses a large area in RAM is required to control the sensitivity for the talker. The AEC suppresses the echo after the directional AGC by using an adaptive filter. The adaptive filter is adapted when the level of the received signal is higher than preset threshold level. The loss controller is used to prevent howling, when sufficient echo suppression is not achieved by the AEC. The hardware is shown in Fig. 2. The unit includes four unidirectional microphones, a loudspeaker, and a DSP board.

5. EXPERIMENTS

The performance of the directional AGC was evaluated by recording data in a real room. The room reverberation time was 300 ms. The arrangement of microphones and sound sources is shown in Fig. 3. There was a hands-free unit including four microphones and a loudspeaker on the table. Other loudspeakers were used as the two talkers and a noise source. The noise source was Hoth noise, and its SNR was 25 dB at the unit. Figure 4 shows the waveform of the signal that was picked up by the microphones and then added. Talkers A and B spoke in turn, where the positions of the talkers were 50 cm and 200 cm from the unit.

5.1. Characteristics of output level

Figure 5 shows the output levels for (a) addition only, (b) conventional AGC, and (c) the proposed directional AGC.

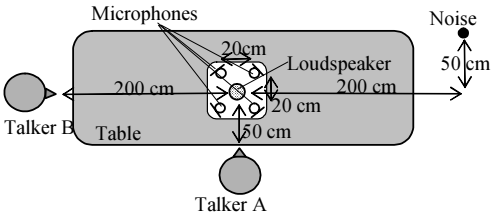

Figure 3: *Arrangement of the microphones and sound sources.*

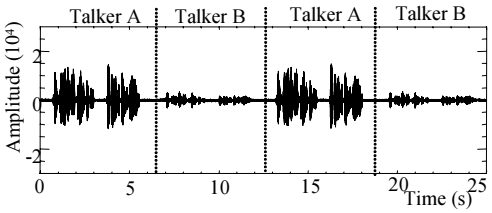

Figure 4: *Signal picked up by the microphones and then added.*

The level of talker B with addition only was 15 dB lower than the appropriate level, because talker B was far from the microphone. In such a case, the participants at the far end often cannot catch the conversation of talker B.

With the conventional AGC, the level of talker B was adjusted to the appropriate level. However, the gain was too large or too small when the talker changed, and it took several seconds to achieve the appropriate level. Furthermore, the noise level increased with the gain adjustment.

The proposed directional AGC kept appropriate output level even when the talker changed because it stored the effective gains for all talkers and controlled the directivity pattern to achieve constant and uniform speech levels for all talkers. Furthermore, the noise level with the proposed AGC was always lower than with addition only. The maximum noise reduction was about 15 dB greater than the conventional AGC because the directional AGC suppressed stationary noise by forming a null point in the directivity pattern.

5.2. Directivity patterns

The directivity patterns of the directional AGC are shown in Fig. 6. In the directivity pattern at point α, the null point was formed toward the noise source, and sensitivity to talker A was kept at 0 dB. Thus, the directional AGC was able to suppress the noise while keeping the sensitivity toward talker A. The patterns at points β and γ show the effect of adjusting the sensitivity while talker B was speaking. After talker B spoke, the sensitivity toward talker B became 13 dB while the sensitivity to talker A was kept. This directivity control enabled the directional AGC to adjust the levels of multiple talkers individually and to suppress the stationary noise.

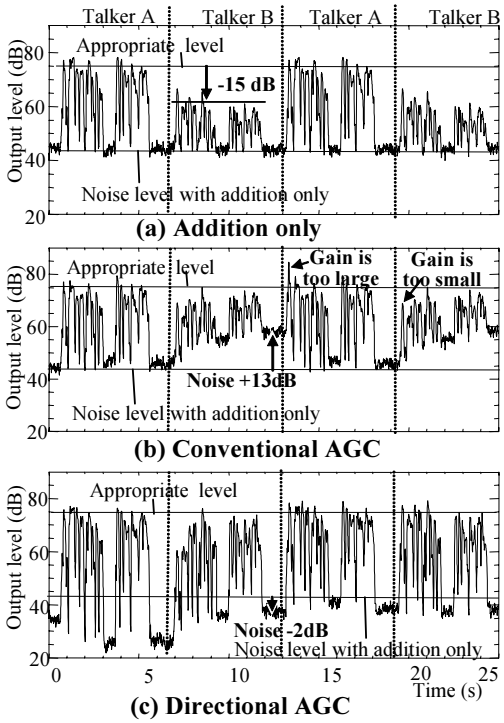

Figure 5: *Output levels with (a) addition only, (b) conventional AGC, and (c) proposed directional AGC.*

5.3. Echo suppression

We evaluated the performance of echo suppression with the directional AGC. To evaluate only the echo suppression achieved by the directional AGC, we observed the echo signal before the AEC while the speech signal was output from the loudspeaker in the unit. Figure 7 shows output levels for the directional AGC (a) not including and (b) including echo suppression. In Fig. 7(a), the echo level is high. In particular, after talker B spoke, the echo level was higher than the speech level of the near-end talker. Thus, in the directional AGC without echo suppression, the directivity control increased the echo signal. Fig. 7(b) shows that the directional AGC with echo suppression suppressed the echo. The echo level was 25 dB lower than without echo suppression. These results show that the directional AGC is effective at suppressing echoes by adding the echo suppression condition in the filter design.

6. CONCLUSIONS

We implemented our previously proposed directional AGC in a hands-free unit. The directional AGC can adjust the speech levels for all talkers in a teleconferencing system. It also suppresses noise and echoes to achieve clear speech. Experiments showed that the unit provides effective gain control for multiple talkers.

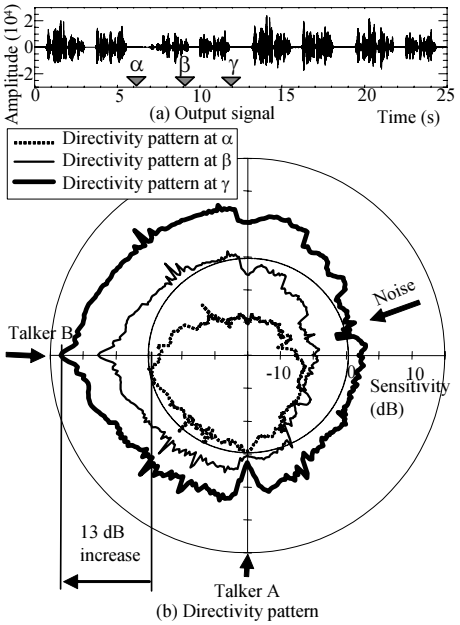

Figure 6: *Directivity patterns of the directional AGC.*

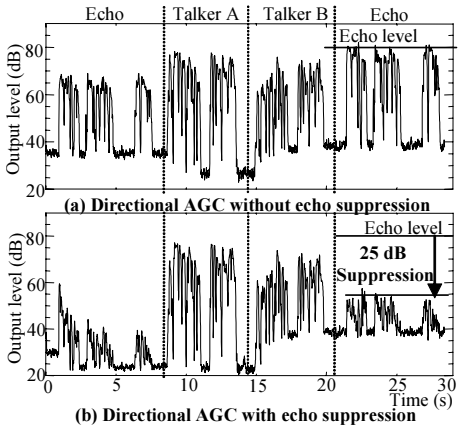

Figure 7: *Performance of the echo suppression.*

7. REFERENCES

[1] K. Kobayashi, K. Furuya, Y. Haneda, and A. Kataoka, "A microphone array system for directional automatic gain control," *IEICE Trans. (A) (in Japanese) Vol. J87-A No. 12*, pp. 1491–1501, 2004.

[2] George R. Steber, "Digital signal processing in automatic gain control," *IECON-88*, pp. 381–384, 1988.

[3] Peter L. Chu, "Voice-activated AGC for teleconferencing," *ICASSP-96, vol. 2*, pp. 929–932, 1996.

[4] Y. Kaneda and J. Ohga, "Adaptive microphone-array system for noise reduction," *IEEE Trans. on ASSP Vol. 34, No. 6*, pp. 1391–1400, 1986.