A New Control Method for an Adaptive Noise Canceller Using Stochastic difference between Voice and Noise Signals Power Change

H. Nishi* and T. Kakinoki*

*Department of Engineering, SOJO University, 860-0844, Kumamoto, Japan
(Tel: +81-96-326-3659; Fax: +81-96-326-3000; Email:nishi@pe.sojo-u.ac.jp)

Abstract: This paper reports a technique for discriminating double talk and echo path change using the stochastic characteristics of power change for an adaptive noise canceller. The causes of rapid error increasing are double talk and echo path change. When the echo path is changed, the system corrects the impulse response in order to reduce the error. However, in the case of double talk, the system has to suspend the updating impulse response in order to maintain the quality of the voice signal. In the conventional system, it was difficult to discriminate between the two situations. In this research, the stochastic characteristics of the voice power change in the double talk period were experimentally verified to be different from the power change during echo path changing. Based on the results, a new double talk detection method is proposed.

Keywords: control, adaptive noise canceller, power change, adaptive filter

1. Introduction

In order to extract a voice signal buried in noise, or to separate sound signals from two kinds of sound sources, adaptive noise cancellers have been studied[1][2][3]. In speech recognition, attention is focused on the pre-processing in order to improve the recognition accuracy.

Adaptive noise cancellers use the technology of adaptive filters with input from two microphones. One is for a reference noise input and the other is for a voice signal where the noise overlaps[4][5].

During the period when there is no voice signal, the noise signal into the voice signal microphone is approximately estimated as the reference noise influenced by the echo path between two microphones using the adaptive filter theory. After the completion of the echo path presumption, it is supposed that the input of a voice signal microphone deducted by the estimated noise signal, which is the convolution between the echo path and the reference noise microphone input, is the voice signal without noise. However, if the same processing as described above is used for a voice signal (double talk period), an incorrect echo path will be estimated and the distortion of the voice signal will increase. For this reason, a method that discriminates whether a voice signal exists or not and suspends the echo path correction during the double talk period was studied.

2. Principle of an adaptive noise canceller and double talk problems

Fig.1 shows the system structure of an adaptive noise canceller. Noise is mainly inputted into a reference noise microphone. It is assumed that the level of the voice signal that is input to the reference noise microphone can be ignored. Therefore, a time series signal of reference noise is expressed with $n(k)$.

On the other hand, the voice signal and noise after being influenced by acoustic space is inputted into the voice microphone. $s(k)$is a mixed time series signal of voice and noise.

$h(k)$ is the impulse response of the acoustic space between

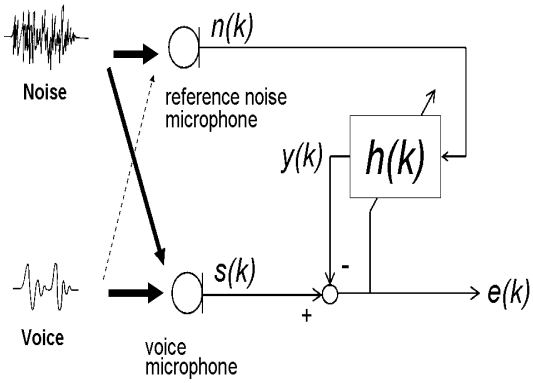

Fig. 1. System structure of an adaptive noise canceller

the reference noise microphone and the voice microphone. Therefore, the estimated value of the noise inputted into the voice microphone is obtained as $y(k)$ which is the convolution of $n(k)$ and $h(k)$.

$$y(k) = \sum_{i=0}^{p-1} n(k-i)h(i) \tag{1}$$

p is the tap length, which means the length of the impulse response. $e(k)$ is the error signal of the noise canceller, which is obtained by subtracting $y(k)$ from $s(k)$. After $h(k)$'s converging, $e(k)$ becomes small enough. Each coefficient of $h(k)$ is updated in order to make the error value $e(k)$ smaller.

$$e(k) = s(k) - \sum_{i=0}^{p-1} n(k-i)h(i) \tag{2}$$

Many different methods for correcting $h(k)$ have been studied[6][7]. In this paper, an LMS(Least Mean Square) algorithm is adopted because the amount of calculation is small and it is used in many systems. The correction formula of the coefficients is described as follows,

$$h_{k+1}(i) = h_k(i) + 2ce(k)n(k-i) \tag{3}$$

c is the step gain. In order to avoid confusing the time and the order number of coefficients, k is used for time and i for

the order number. Fig.2 shows the waveform examples of s(k) and e(k). Waveform (a) is an example of s(k). n(k)

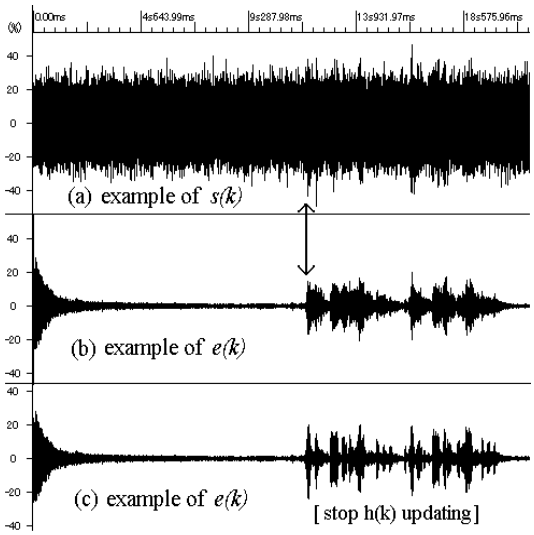

Fig. 2. Waveform examples of s(k) and e(k)

is white noise and the S/N ratio of s(k) is approximately 3dB-0dB. Waveform (b) shows e(k). The arrow in (a) and (b) indicates the beginning point of the double talk period. The S/N ratio of e(k) is secured more than 20dB and noise is sufficiently eliminated. However, several unnatural envelopes are observed after power peaks in the waveform (b). Consequently, an excessive echo is added.

Voice is a desirable signal for users, but it is an error signal for the system. The noise canceller feedbacks wrong data to h(k) in order to reduce the error(voice signal). e(k) waveform in which the double talk period is given manually and updating is stopped in the double talk period is shown in (c) of Fig.2. In waveform (c), the natural start and stop of the voice signal can be observed. Moreover, the echo is limited and a natural voice can be obtained. The results obtained from a survey regarding the voice quality for both data in which echo path updating is suspended while double talk like(c) and data without double talk detection like(b) is shown in Fig.3.

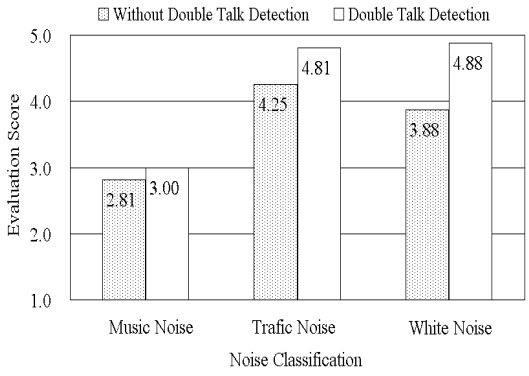

Fig. 3. Opinion test result for voice quality

Regardless of the noise category, the evaluation results show that the data obtained with suspending echo path updating during double talk is better than the data obtained without double talk detection. In terms of a practical system, the double talk must be detected automatically not manually. Although many double talk detection methods have been studied, no perfect system has been reported. Most methods aim at the power or correlation of n(k), s(k) or e(k) and detect the double talk when the value exceeds the threshold. On the other hand, the waveform of s(k) and e(k) when the echo path changes shown by Fig.4 are similar to that of the double talk period. Then it is difficult to distinguish the double talk and the change of the echo path.

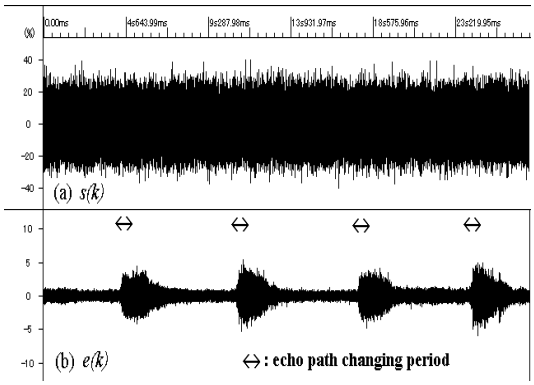

Fig. 4. Waveform of s(k) and e(k) when echo path changes(echo path changes four times-indicated by arrows)

3. Conventional researches

Adaptive noise cancellers and acoustic echo cancellers have been studied on the same theoretical basis - adaptive filter theory. Many conventional researches for double talk detection have been applied for acoustic echo cancellers. In the case of acoustic echo cancellers, n(k) in Fig.1 is the voice signal of tele-conference users. However, n(k) is the reference noise input in adaptive noise cancellers.

The conventional researches are classified into time domain processing[8] and frequency domain processing[9]. Time domain processing uses duration or power level. Frequency domain processing observes correlation coefficients in order to detect double talk. In the case of the duration, by using the stochastic characteristic of human voice duration, the system can detect the double talk when extraordinary duration is observed. In the case of correlation, it is possible by observing smaller correlation coefficients between n(k) and s(k) than the threshold value. However, those techniques are not successfully for adaptive noise cancellers because n(k) is high level noise especially continuous noise. Apparently duration is not useful and correlation is not effective when the power level of n(k) is very high.

Such situation means other parameters should be studied. In this research, the power change of the voice signal in the double talk period is experimentally verified to be different from the power change during echo path changing. In addition, a method of discriminating the double talk and the echo path change is proposed. First of all, a voice signal and the remaining error signals are collected, and the histograms of the power change value between frames are obtained. Using the histograms, the voice signal probability (a posteriori proba-

bility) of the data whose category is unknown and the given power change value are studied. The voice signal probability means the double talk probability. The double talk detection method is proposed in this case because it judges the data and determines that the beginning frame of the voice is where a posteriori probability exceeds the fixed threshold. The accuracy of this double talk detection method is evaluated compared with the conventional method, which detects the double talk by using only the power level.

4. Power change in the double talk and echo path changing period

Discriminating the double talk period and the echo path changing period was shown to be important in the previous section. In this section, the following conditions are studied in order to attain the goal.

(1) Power level of e(k) is small enough while double talk does not exit after the converging of h(k).

(2) Causes of echo path change are the moving of the noise source, microphone and speaking persons. Then the changing rate is low(100ms-1s).

(3) Power change of the voice signal depends on the speech speed(20ms-100ms), then it is faster than the case of echo path change.

Therefore, the period where only a reference noise exists can be detected using the power level threshold under condition (1), and it is expected that echo path changing and double talk are discriminated using the threshold for power change under condition (2) and (3). This section studies the stochastic characteristics of power change while changing the period of the echo path and the double talk period, then verifies the possibility of discrimination between them[10].

4.1. Experimental conditions for training data collection

The measurement conditions of power change in the double talk period and the echo path changing period are shown in Table 1. Power data was measured in a room under the conditions described in Fig. 5. Two audio signals from a reference noise microphone and a voice microphone were recorded simultaneously using stereo recording equipment. When the echo path must be changed, the shield screen is moved.

Table 1. Measurement conditions of power change

Items	Contents
Noise type and data length (n(k))	·White noise: 32.5s ·Traffic noise: 43.0s ·Music noise: 37.6s
Double talk data	Voice and white noise Data length: 79.1s
Method of echo path changing	Moving a shield screen near voice microphone
Sampling frequency	11.025kHz
Quantization	16bit
Frame length	18.14ms(200samples)
Measured data	Power and Power change (dB) of 3 kinds of noise and voice data in the error signal e(k)
Calculation method	LMS algorithm

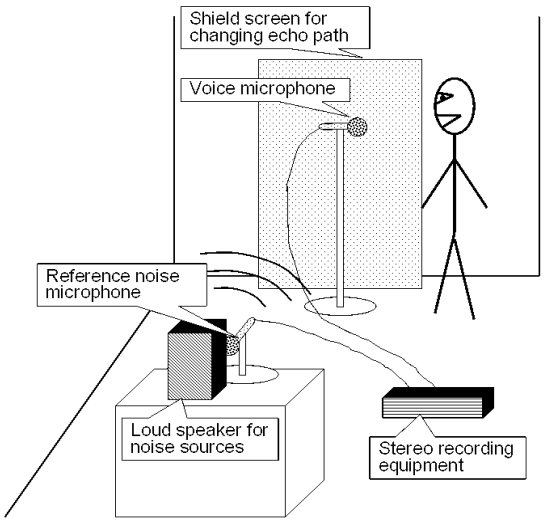

Fig. 5. Experiment environment

4.2. Results of the experiment

Fig.6 shows the histogram of power change of the noise and voice data. Each data represents a difference in the power value between two bounded frames. The average and the standard deviation are shown in Table 2. The average of each category is close to 0. However, the standard deviation is different for each category. The standard deviation of voice is the largest. This means the power change of the beginning

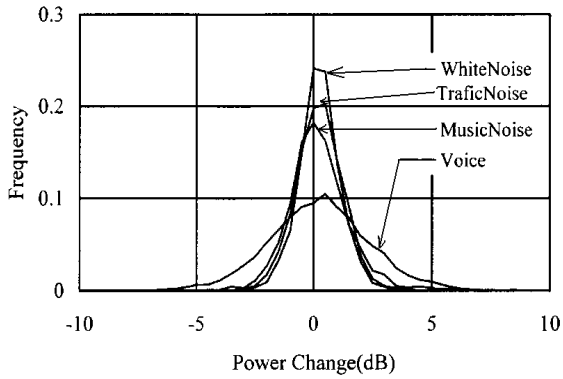

Fig. 6. Histogram of power change

Table 2. Average and standard deviation of each category

Category	Average	Standard deviation
White noise	-0.00973	0.8968
Traffic noise	-0.01350	1.0167
Music noise	-0.01064	1.3697
Voice	0.00114	2.1702

point and the ending point of voice is larger than that of any other category. In three kinds of noise data, the standard deviation of music noise is larger than that of the other two categories. White noise is the smallest. The S/N ratio for any category is close, therefore the stochastic characteristics of each category differs as shown.

4.3. Analysis of the power change data

In this section, when the category is unknown and the power change is given, a posteriori probability that the category is voice(means double talk) is introduced. The following conditions are supposed.

(a) Echo path($h(k)$) is converged already.
(b) Error noise is small.
(c) In this situation, the double talk or echo path change is detected using the power level.
(d) The a priori probabilities of double talk and echo path change are the same.

According to item(c), the system should judge whether a double talk or echo path change has begun. Here, several variables are defined.

E_{dt} Event where double talk appears.
E_{ep} Event where an echo path change appears.
O_{pc} pc is power change data which means the power difference between the frame and the previous frame. O_{pc} is the event where the power change data is pc.

Simultaneous probability of event E_{dt} and O_{pc} is expressed as the formula,

$$\begin{aligned} Pr(E_{dt} \cap O_{pc}) &= Pr(O_{pc}|E_{dt})Pr(E_{dt}) \\ &= Pr(E_{dt}|O_{pc})Pr(O_{pc}) \end{aligned} \quad (4)$$

Using the Bayesian rule,

$$Pr(E_{dt}|O_{pc}) = \frac{Pr(O_{pc}|E_{dt})Pr(E_{dt})}{Pr(O_{pc})} \quad (5)$$

The denominator is rewritten as following,

$$Pr(E_{dt}|O_{pc}) = \frac{Pr(O_{pc}|E_{dt})Pr(E_{dt})}{Pr(O_{pc}|E_{dt})Pr(E_{dt}) + Pr(O_{pc}|E_{ep})Pr(E_{ep})} \quad (6)$$

A priori probabilities are the same under condition (d),

$$Pr(E_{dt}|O_{pc}) = \frac{Pr(O_{pc}|E_{dt})}{Pr(O_{pc}|E_{dt}) + Pr(O_{pc}|E_{ep})} \quad (7)$$

$Pr(O_{pc}|E_{dt})$ and $Pr(O_{pc}|E_{ep})$ can be obtained by Fig.6. Therefore, $Pr(E_{dt}|O_{pc})$ is calculated using the formula(6) and Fig.6. The result of the calculation is shown in Fig.7.

The calculated threshold values of the power change using Fig.7 are shown as Table.3. At the same accuracy(a posteriori probability), for example 99%, the beginning point of the double talk can be detected by observing the 3.4dB power change in the case of voice with white noise. However, in the case of music noise, it is necessary to detect 5.9dB power change. As a result of the difference, misdetection increases for voice with music noise. In the case of voice with traffic noise, double talk detection accuracy is the middle between white noise and music noise.

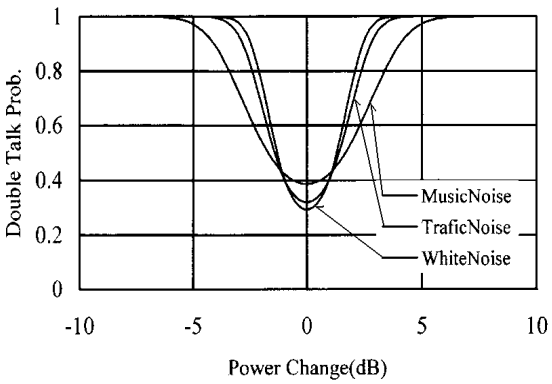

Fig. 7. A posteriori probability of double talk

Table 3. Threshold of power change value for detecting the beginning point of double talk(dB)

Category	A posteriori probability		
	90%	95%	99%
White noise	2.4	2.9	3.4
Traffic noise	2.7	3.2	3.9
Music noise	4.1	4.6	5.9

5. Double talk detection method

In the previous section, the stochastic basis in order to obtain the double talk detection method is described. In this section, the specific method is introduced.

Double talk detection is composed of two modules. One is to detect the beginning point of the double talk and the other is to detect the end point of the double talk. The theoretical ground behind detecting the beginning point has already been discussed. In the case of detecting the beginning point, before beginning double talk, the power level is low enough, therefore, the threshold value of the power change can be set according to the method described in the previous discussion.

On the other hand, while searching the ending point, the state is in a double talk period itself, then, the power level is high and unstable. In such a situation the power change information is not useful.

In order to solve the problem, a method that uses different data between the detection of the beginning point and the ending point of the double talk is proposed. For the beginning point, the power change is used according to the discussed method, and for the ending point, the threshold power level is obtained using the data before the beginning point is introduced.

The threshold for the ending point is set as 3dB plus the lowest value among the 50 frames before the beginning point. A common value is used as the threshold for detecting the beginning point. It is set using a desirable a posteriori probability and the Fig.7. On the other hand, the ending point is established dynamically. Furthermore, the threshold for ending point detection is determined using voice data before

the beginning point. This is done to avoid setting an abnormally large value when an unexpectedly long voice exists. However, after stop the updating the adaptive filter by detecting the beginning point, if the end point cannot be detected, the situation produces the problem that system cannot restart the updating the adaptive filter. In order to avoid the problem, when the end point was not detected within the fixed time, the new method restarts the updating automatically.

6. Evaluation of the new double talk detection method

In this section, the new double talk detection method proposed in the previous section is evaluated. The evaluation experiment consists of two items.

(i) Accuracy of double talk detection Probability that the double talk is detected correctly when double talk occurs.

(ii) Accuracy of echo path change detection Probability that the echo path change is detected correctly when an echo path change occurs.

The experiment environment is the same as that of the training data collection shown in Fig.5.

6.1. Accuracy of double talk detection

In this section, the accuracy of double talk detection is evaluated in the case where the evaluation data does not contain an echo path change. The evaluation result is shown in Table 4 and Fig.8. The evaluation data is open and different from the training data collected in section4.

The detection accuracy in Fig.8 is obtained using the observed data and the following calculation.

$$\mathrm{DetectionAccuracy} = \frac{\mathrm{CorrectlyDetectedFramesNumber}}{\mathrm{TotalFramesNumber}}$$

This evaluation data has no echo path change. It has a noise period and a double talk period. Then this evaluation represents the detection accuracy when the double talk occurs.

Table 4. Result of double talk detection accuracy (Frame length is 18.14ms)

Items	Noise Category		
	White noise	Traffic noise	Music
Data length(s)	54.50	48.10	48.09
Frame number	3005	2652	2651
Correct frame number	2809	2332	2258
Accuracy(%)	93.5	88.0	85.2

The accuracy obtained in white noise is the best, in traffic noise is the next best and in music noise is the worst. This result is also expected using the data of Fig.7. However, this accuracy is almost the same as that of the conventional method (using the power level) because the updating of the echo path is converged and the background noise is small enough. This means that the next experiment, which detects the echo path change, has become more important.

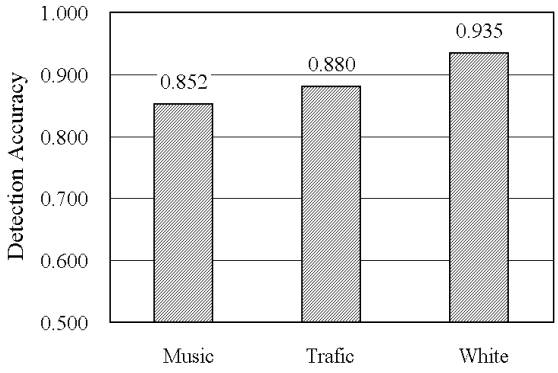

Fig. 8. Accuracy of double talk detection

6.2. Accuracy of echo path change detection

This section describes the detection accuracy when the echo path changes. The evaluation result is shown in Table 5 and Fig.9. The evaluation data is also open and different from the training data collected in section4. The detection error rate in Fig.9 is obtained using the observed data and the following calculation.

Table 5. Result of echo path change detection accuracy

Items	Noise Category		
	White noise	Traffic noise	Music
Data length(s)	100.77	63.22	103.36
Frame number	5555	3485	5698
Error rate of Conventional method(%)	22.6 (1250)	30.0 (1042)	52.2 (2984)
Error rate of new method(%) (frame number)	2.8 (135)	5.0 (171)	17.2 (990)

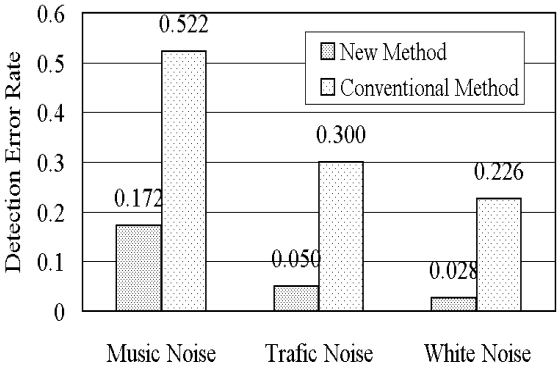

Fig. 9. Detection error of echo path change

$$\mathrm{DetectionErrorRate} = \frac{\mathrm{ErrorFramesNumber}}{\mathrm{TotalFramesNumber}}$$

This evaluation data has no double talk. It has a noise period and an echo path changing period. Therefore this evaluation indicates the error rate when the echo path change occurs.

Among the three noise categories, the new method is better than the conventional method, which uses only the power level. It was proven that the idea of using the stochastic characteristics of power change is useful and robust. However, the detection ability is not stable. It depends on the noise category. In the case of white noise and the traffic noise, the detection error of echo path changing, is small. However, in the case of music noise, the difference between the new method and the conventional method is not sufficiently large. The reason is that the power change in music is large compared to that of the other two kinds of noise. It is a future subject to be investigated in order to improve the double talk detection ability in music noise.

7. Conclusion

This paper proposes a new double talk detection method of distinguishing from an echo path change using the power change between two adjacent frames. The method is evaluated using mixed signals, which include voice and three kinds of noise; white noise, traffic noise and music noise. The evaluated items are the accuracy of double talk detection and the detection error of the echo path change. The new method was confirmed to be better than the conventional method in the three noise categories.

References

[1] B. Widrow, J. R. Glover, "Adaptive noise canceling: principles and applications," Proc. IEEE, 63, 12, pp.1692-1716, 1975.

[2] W. A. Harrison, S. J. Lim, E. Singer, " A new application of adaptive noise cancellation," IEEE Trans. on Acoustics, Speech and Signal Processing, ASSP-34, pp. 21-27, 1986.

[3] Liem M, Manck O," Architecture of a Single Chip Acoustic Echo and Noise Canceller Using Cross Spectral Estimation," Proc ICASSP, Vol.2003, No.Vol.2, PageII.637-II.640, 2003

[4] B. Widrow, S. Sterns, " Adaptive Signal Processing," Prentice-Hall, 1985.

[5] Simon Haykin, " Adaptive Filter Theory Third Edition," Prentice Hall, 1996.

[6] J. Nagumo, A. Noda, " A learning method for system identification," IEEE Trans. on Automat. Control, Vol. AC-12, no. 3, pp.282-287, June 1967.

[7] T. Usagawa, H. Matsuo, Y. Morita, M. Ebita, " A New Adaptive Algorithm Focused on the Convergence Characteristics by Colored Input Signal: Variable Tap Length LMS," IEICE Trans. Fundamentals, Vol. E75-A, 11, 1992.

[8] S. Ninami and T. Kawasaki, " A Double Talk Detection Method for an Echo Canceller", ICC'85 REC., pp.1492-1497, 1985.

[9] T. Gaensler and J. Benesty, " A Frequency-Domain Double-Talk Detector Based on a Normalized Cross-Correlation Vector", Signal Process, Vol.81, No.8, pp.1783-1787, 2001.

[10] H. Nishi and M. Kitai, " Analysis and Detection of Double Talk in Telephone Dialogs," Proc. ICSLP94, S27-11.1, 1623-1626, 1994.