

# Noise Suppression Signal Processing Using 2-Point Received Signals

Yutaka Kaneda and Mikio Tohyama, Regular Members

Musashino Electrical Communication Laboratory, N.T.T., Musashino, Japan 180

## SUMMARY

A number of research reports has been presented on the noise suppression signal processings as one of the important technical problems. Although a reduction feeling of the noise level can be obtained by the conventional method using the one-point received signal, it has been pointed out that the program of improving the articulation is not sufficient. To solve this problem, we propose a noise suppression signal processing method using two-point received signals. As the first step, the noise suppression method is discussed assuming that the desired signal is uncorrelated to the noise. It is shown that the noise suppression filter, which is the optimum in the sense of least-mean-square error, is obtained from the correlation coefficients between the frequency components of the two-point received signals. Then a method is shown in which the two-point receiving noise suppression system can be constructed using FET based on the proposed principle. Finally, to verify the effectiveness of the proposed processing, the method is applied to the two-point received speech under room noise. The articulation tests were performed for the processed and unprocessed speeches. As a result, it was seen that the syllable articulation score could be improved by 4 to 10 percent when the correlation coefficient between the two-point received signals of the noise is 0.1 or less and that of the speech is 0.9 or more.

## 1. Introduction

For the received signal, which consists of speech plus noise components, the signal processing to suppress the noise component is a technical problem requested in many practical situations. Examples are

telephone communication under high noise level circumstances such as railroad station, busy street and within an automobile, and loud-speaking telephone system even under low noise level. This paper considers the noise suppression assuming acoustic noises.

With the recent development of digital technology, there have emerged various real-time processings as possible means to solve this problem. Based on those technologies, a number of researches has been made on the signal processing to suppress the noise. Typical methods are spectrum subtraction [1], Wiener filtering [2], comb filtering [3], SPAC [4], and adaptive filtering [5]. By those methods, subjective noise reduction is achieved, while it is pointed out that the phoneme articulation is degraded in most of the methods [6]. In addition, a priori knowledge is required in the processing for the characteristics of the desired signal (such as speech and musical sound) or waveform or power spectrum of the noise, which is a restriction in the application. In contrast to these past methods, which receive the signal sound at one point, we propose the new noise suppression technique using two-point received signals aiming at the solution of the problem by utilizing the spatial property of the sound field.

As the processing by two-point receiving, Allen's reverberation suppression technique is known [7]. In this method, the two-point receiving is used for the signal with added room reverberation, and the reverberation suppression filtering is made utilizing the fact that the correlation between the two-point received reverberation is small compared with that of direct sound and the early echoes. There are some problems left unsolved in this method, such as optimization of the filter coefficients. On

the other hand, it is known that when the two-point receiving of the noise in the sound field is made, the correlation between the two received noises is almost zero, if the two receiving points are sufficiently far apart [8]. From such a viewpoint, we investigated the noise suppression filtering considering the correlation between the two-point received signals.

In this paper, Sect. 2 formulates the noise suppression problem, and it is shown through analysis that the optimum filter coefficients can be calculated from the two-point received signals. The result is the same as the filter coefficients obtained by the conventional Wiener filter. Although the filter coefficients are determined by the long-term average spectrum of the noise, which must be obtained as an a priori information in the conventional Wiener filtering method, the proposed method is advantageous in that the filter coefficients are determined directly from the signal containing both the desired signal and the noise. Then, in Sect. 3, the procedure for the processing by this method is described. Finally, in Sect. 4, the experimental result of subjective evaluation is described for the quantitative evaluation of the noise suppression by this proposed method.

## 2. Processing Principle

It is shown in this section that the optimum noise suppression filtering (in the sense of the least-mean-square error) can be realized by calculating the correlation coefficient  $R_k$  between the corresponding frequency components of the two-point received signals. A discrete time-series signal is considered as the signal.

### 2.1 Optimum noise suppression filter

Assume that the signal  $x(n)$  consists of the desired signal  $s(n)$  and the noise  $n_1(n)$ , and it is represented as

$$x(n) = s(n) + n_1(n) \quad (1)$$

The signal is considered as a discrete sequence, and  $n$  is the parameter representing time.

$x(n)$ ,  $s(n)$  and  $n_1(n)$  are then decomposed into independent contiguous  $N$  frequency components, which are denoted by  $x_k(n)$ ,  $s_k(n)$  and  $n_{1k}(n)$  ( $k = 1, 2, 3, \dots, N$ ). Then,  $x(n)$  is represented as follows:

$$x(n) = \sum_{k=1}^N x_k(n) = \sum_{k=1}^N (s_k(n) + n_{1k}(n)) \quad (2)$$

Consider the noise suppression filtering, in which the frequency component  $x_k(n)$

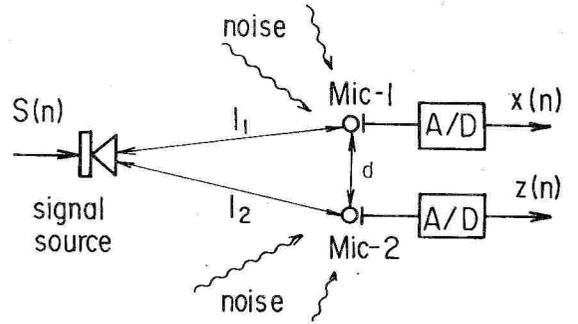


Fig. 1. Schematic diagram of 2-point receiving.

of  $x(n)$  is weighted by constant (filter coefficients)  $c_k$  to produce the output  $y(n)$

$$y(n) = \sum_{k=1}^N c_k \cdot x_k(n) \quad (3)$$

and consider the evaluation function  $I$  representing the mean squared error

$$I = \overline{|y(n) - s(n)|^2} \quad (4)$$

Therefore, the filter coefficients  $\tilde{c}_k$  minimizing  $I$  gives the optimum filter. Above, the superline "-" indicates the time averaging, which is defined for an arbitrary function  $g(n)$  as

$$\overline{g(n)} \triangleq \lim_{M \rightarrow \infty} \frac{1}{2M+1} \sum_{n=-M}^M g(n) \quad (5)$$

Using Eq. (3) and independent property of frequency components, Eq. (4) can be written as

$$\begin{aligned} I &= \overline{\left| \sum_{k=1}^N \{c_k x_k(n) - s_k(n)\} \right|^2} \\ &= \sum_{k=1}^N \overline{\{c_k^2 x_k^2(n) - 2c_k x_k(n) s_k(n) + s_k^2(n)\}} \end{aligned} \quad (6)$$

Assuming that the desired signal  $s(n)$  and the noise  $n_1(n)$  are uncorrelated, and substituting Eq. (2) into  $x_k(n)$  of Eq. (6), Eq. (6) turns out to be

$$I = \sum_{k=1}^N \overline{\{c_k^2 (s_k^2(n) + n_{1k}^2(n)) - 2c_k \cdot s_k^2(n) + s_k^2(n)\}} \quad (7)$$

Forming the partial derivatives of Eq. (7) in regard to  $c_k$  and setting the derivatives as zero, the optimum filter coefficients  $\tilde{c}_k$  minimizing  $I$  are obtained as

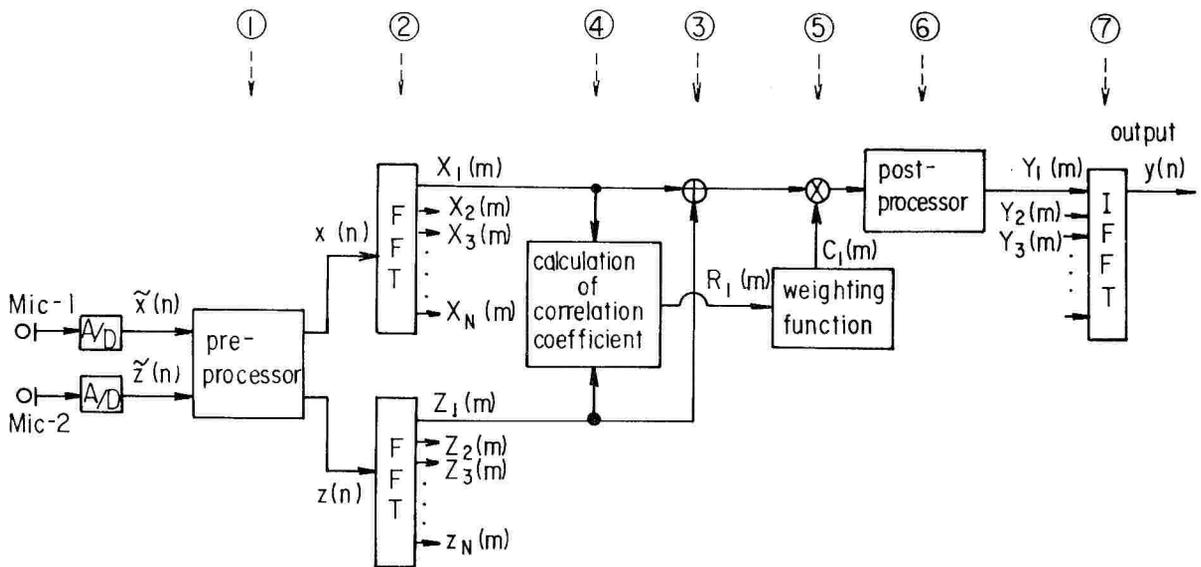


Fig. 2. Block diagram of processing.

$$\tilde{c}_k = \frac{\overline{s_k^2(n)}}{\overline{s_k^2(n) + n_{1k}^2(n)}} \quad (8)$$

## 2.2 Noise suppression by two-point receiving

The optimum filter coefficient  $\tilde{c}_k$  described in the previous section cannot be obtained from the one-point received signal containing both the desired signal and the noise. Consequently, in this method, the desired signal is received at two points and the optimum filter coefficients are determined by the following method.

Figure 1 shows the two-point receiving of the desired signal  $s(n)$  in a sound field containing noise. It is assumed that the reflected sounds in the room have sufficiently low levels compared with the desired signal. To simplify the description, it is assumed that the distances  $l_1$  and  $l_2$  between the receiving points and the signal source are equal ( $l_1 = l_2$ ). The propagation delay for the desired signal is ignored.

Then the two signals  $x(n)$  and  $z(n)$  received at two points are represented as follows:

$$x(n) = s(n) + n_1(n) \quad (9)$$

$$z(n) = s(n) + n_2(n) \quad (10)$$

where  $n_1(n)$  and  $n_2(n)$  are the noise components contained in the signals received by the two microphones.

For the noises received at two points, the correlation coefficient between the corresponding frequency components is in general small, if the distance  $d$  between the two receiving points is large. In particular, in the diffused sound field, it converges to zero [8]. Consequently, the following two assumptions are made for the frequency components  $n_{1k}(n)$  and  $n_{2k}(n)$  of the two-point received noises.

Assumption 1.  $s_k(n)$ ,  $n_{1k}(n)$  and  $n_{2k}(n)$  are uncorrelated to each other.

Assumption 2. The powers of  $n_{1k}(n)$  and  $n_{2k}(n)$  are equal. In other words,

$$\overline{n_{1k}^2(n)} = \overline{n_{2k}^2(n)} \quad (11)$$

Under those assumptions, the correlation coefficient  $R_k$  between the frequency components  $x_k(n)$  and  $z_k(n)$  of the two-point received signals is calculated:

$$R_k = \frac{\overline{x_k(n) \cdot z_k(n)}}{\{\overline{x_k^2(n)} \cdot \overline{z_k^2(n)}\}^{1/2}} \quad (12)$$

Substituting Eqs. (9) and (10),

$$\begin{aligned} R_k &= \frac{(s_k(n) + n_{1k}(n))(s_k(n) + n_{2k}(n))}{\{(s_k(n) + n_{1k}(n))^2 \cdot (s_k(n) + n_{2k}(n))^2\}^{1/2}} \\ &= \frac{\overline{s_k^2(n)}}{\overline{s_k^2(n) + n_{1k}^2(n)}} \end{aligned} \quad (13)$$

which is the same as the value of  $\tilde{c}_k$  represented by Eq. (8).

Consequently, the optimum filter in the sense of the least-mean-square error is realized by using the correlation coefficient between the corresponding frequency components of the two-point received signals as the filter coefficient for that frequency.

### 3. Signal Processing Procedure

The filtering method described in the previous section is to determine the ratio of the desired signal components in each frequency component through a long-term observation and to perform the filtering using the ratio as the coefficient. When the speech is considered as the desired signal, the short-term spectrum of the speech signal varies with time, and the short-term spectrum has the important phonetic information. Consequently, as the means of noise suppression for the speech signal, the adaptive filtering will be more effective than the time-invariant filtering described in the previous section. In the former, the method of the previous section is applied to the signal within a certain interval and the filtering coefficients are changed according to the short-term spectrum of the speech signal. We constructed the following noise suppression system utilizing the short-term Fourier transformation using FFT.

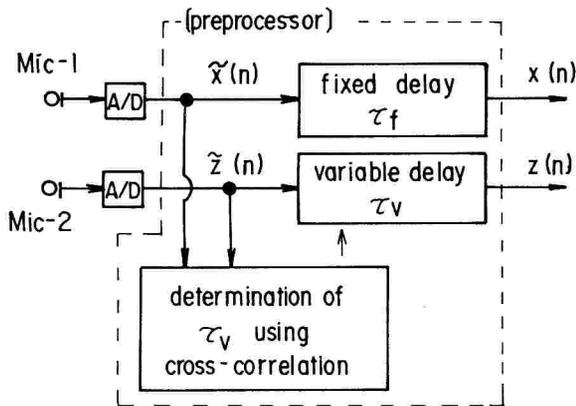


Fig. 3. Preprocessor.

The steps of the procedure and described in the following, based on the schematic diagram of Fig. 2.

(1) Pre-processing: Let the distances from the signal source to the two microphones be  $l_1$  and  $l_2$ . Then the signals  $\tilde{x}(n)$  and  $\tilde{z}(n)$  at the two microphones are received with the time difference

$$\tau_0 = \frac{l_2 - l_1}{c} \quad (14)$$

where  $c$  is the sound velocity.

To adjust the time of the desired signal components contained in the two-point received signals to in-phase, as is indicated by Eqs. (9) and (10), the delay equalization of Fig. 3 is provided. In this processing, the cross-correlation  $\rho_{xz}(\tau)$  between  $\tilde{x}(n)$  and  $\tilde{z}(n)$  is calculated, and the value of  $\tau$  giving the maximum of  $\rho_{xz}(\tau)$  is used as the estimation  $\hat{\tau}_0$  for  $\tau_0$  in Eq. (14).

Then the fixed delay  $\tau_f$  is given to the input  $\hat{x}(n)$ , and the following variable delay is given to  $\tilde{z}(n)$ :

$$\tau_v = \tau_f - \hat{\tau}_0 \quad (15)$$

This procedure compensates for the desired signal time difference that existed between the two inputs to the microphones.

### (2) Partition of frequency band

The frequency band is partitioned by FFT. It is known that when Hamming window is used as the time-window for FFT, if FFT is performed by shifting the window by  $T_F/4$ , where  $T_F$  is the length of the time-window, a satisfactory waveform can be restored [9].

As a result of such an FFT, the short-term Fourier transform  $X_k(m)$  and  $Z_k(m)$  are obtained for each  $T_F/4$  in time:

$$X_k(m) = \sum_{n=-\infty}^{\infty} w_F(m-n) \cdot x(n) \cdot e^{-j\omega_k n} \quad (16)$$

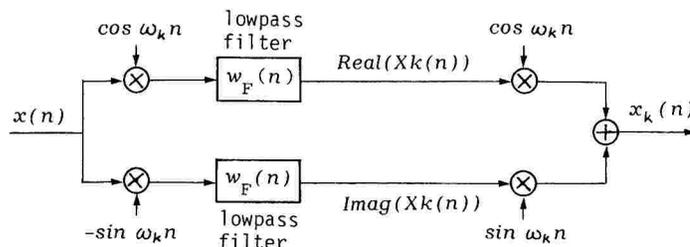


Fig. 4. Relationship between short-term Fourier transform and frequency component.

$$Z_k(m) = \sum_{n=-\infty}^{\infty} w_F(m-n) \cdot z(n) \cdot e^{-j\omega_k n} \quad (17)$$

where  $w_F$  is Hamming window,

$$\omega_k = 2\pi(k-1)/T_F \quad k=1, 2, \dots, N$$

and  $m$  is the time parameter at interval of  $T_F/4$ .

In the following processings (3) to (6), the same processing is made on each frequency component obtained by the above procedure. Only the processing for  $k=1$  is shown in Fig. 2.

### (3) Noise suppression by addition

The corresponding short-term Fourier spectra  $X_k(m)$  and  $Z_k(m)$  of the two channels are added. Since the desired signals are added in-phase, this processing suppresses the noise component relatively.

### (4) Calculation of correlation coefficient

It is known that the relation between the short-term Fourier spectrum and the frequency component of the signal can be represented as in Fig. 4 [9]. By this relation, the frequency components  $x_k(n)$  and  $z_k(n)$  of the two input signals can be represented as follows, using their Fourier transforms  $X_k(n)$  and  $Z_k(n)$ , respectively:

$$\left. \begin{aligned} x_k(n) &= \operatorname{Re}(X_k(n)) \cdot \cos \omega_k n \\ &\quad + \operatorname{Im}(X_k(n)) \cdot \sin \omega_k n \\ z_k(n) &= \operatorname{Re}(Z_k(n)) \cdot \cos \omega_k n \\ &\quad + \operatorname{Im}(Z_k(n)) \cdot \sin \omega_k n \end{aligned} \right\} \quad (18)$$

Substituting Eq. (18) into Eq. (12), and calculating by assuming that the variations of  $X_k(n)$  and  $Z_k(n)$  are sufficiently slow compared with that of  $\sin \omega_k n$ , the correlation coefficient can be expressed approximately as

$$R_k \cong \frac{\operatorname{Re}(X_k^*(n) \cdot Z_k(n))}{\{|X_k(n)|^2 \cdot |Z_k(n)|^2\}^{1/2}} \quad (19)$$

where  $*$  indicates the complex conjugate.

Based on this expression, we defined the correlation coefficient between the frequency components in the proposed procedure as follows:

$$R_k(m) = \frac{\operatorname{Re}(X_k^*(m) \cdot Z_k(m))^w}{\{|X_k(m)|^{2w} \cdot |Z_k(m)|^{2w}\}^{1/2}} \quad (20)$$

where  $\overline{w}$  indicates the weighted mean-in-time for a finite interval. Letting the averaging window be  $w_R(m)$ , the mean is defined as follows for any function  $g(m)$ :

$$\overline{g(m)}^w \triangleq \sum_{l=-\infty}^{\infty} w_R(m-l) \cdot g(l) \quad (21)$$

### (5) Filtering coefficient and filtering

If the noise components in the two channels are uncorrelated, the correlation coefficient defined by Eq. (12) cannot go negative. In the actual processing, however, an averaging window  $w_R(m)$  for a finite interval is used, and there can be a case where the correlation coefficient  $R_k(m)$  calculated by Eq. (20) is negative for the band containing only noise.

Consequently, the following weighting function  $f(\cdot)$  is used to determine the filtering coefficient  $c_k(m)$ :

$$c_k(m) = f(R_k(m)) \triangleq \begin{cases} R_k(m) & R_k(m) \geq 0 \\ 0 & R_k(m) < 0 \end{cases} \quad (22)$$

The coefficient  $c_k(m)$  is multiplied with the previous result of addition  $X_k(m) + Z_k(m)$ .

### (6) Post-processing (Low-frequency suppression filtering)

The low-frequency component has longer wavelength. Consequently, even if a noise component coming from random direction is received, a large correlation can be produced when the distance between the receiving points is not sufficiently large [8]. Consequently, the post-filtering is used to suppress the low-frequency components.

### (7) Synthesis of time-waveform

The short-term spectra  $Y_1(m)$  to  $Y_N(m)$  obtained from above processings (3) to (6) are used to perform the inverse Fourier transformation, and the result of processing is given by overlapping add method.

## 4. Noise Suppression Experiments by the Proposed Method

To verify the effectiveness of the noise suppression by proposed method, the improvement of articulation was examined. The method and the result of experiment are described in the following.

### 4.1 Sound field for experiments

The experiment was performed in a room with volume 100 m<sup>3</sup>, surface area 160 m<sup>2</sup>,

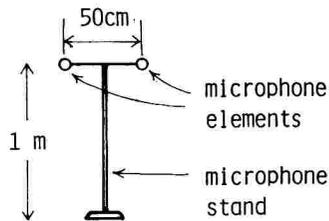


Fig. 5. The arrangement of microphones.

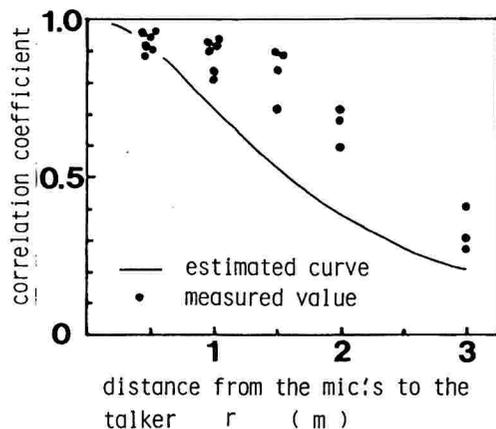


Fig. 6. Correlation coefficients between voice signals received by 2-microphones.

reverberation time 0,4 s, and room-constant 45 m<sup>2</sup>. Those conditions are nearly the same as in an ordinary office or meeting rooms [13]. The noise is due largely to the computer located in the room, with the sound level approximately 65 dB. As the microphones, electret-condenser microphones with diameter 1 cm are located 1 m above the floor with the horizontal distance of 50 cm, as is shown in Fig. 5.

In the processing principle described in Sect. 2, it is assumed that the two-point correlation coefficient of the speech is 1. In the actual field in the room, the correlation coefficient is lowered due to the effect of the reflected sounds. Consequently, based on the approximate formula regarding all reflected sounds superposed on the direct sound as diffused sounds [12], the correlation coefficient between two-point received speech signals in the room is calculated. The result is shown in Fig. 6 by a solid-line with the distance  $r$  between the speaker and the microphones as the parameter. Here, the distance  $r$  between the speaker and the microphones is defined as the distance from the speaker to the mid-point of the two microphones.

Then the correlation coefficients between two-point received speech signals in this room were measured. The condition of the measurement was as follows. The speaker and the microphones are at a distance of 1 m or more from the side wall. The noise source was stopped during measurement. The frequency band to consider is set as 300 to 4000 Hz. The correlation coefficients of the speech signal were calculated for the two-point received signals after delay equalization of Fig. 3.

The correlation coefficients measured for various distances between the speaker and the microphones is shown by dots in Fig. 6. It is seen from Fig. 6 that as the distance  $r$  between the speaker and the microphones is increased, the correlation coefficient is lowered due to the effect of the reflected sounds in the room. In this approximation, all reflected sounds are regarded as diffused sounds with small correlation coefficient. Consequently, the calculated estimation curve gives the lower bounds for the average correlation coefficient of the two-point received speech signals under the given room condition. This is seen from the fact that the measured value is larger than the calculated value except for the case of  $r = 0.5$  m, where the effect of the reflected sounds is very small.

The correlation coefficient of the noise is difficult to estimate theoretically since there is more than one noise source with various dimensions. It is anticipated, however, that the sound field is diffusive with small correlation coefficient between two-point received noises, since there are a number of noise sources and the directivities of noise sources are not remarkable. To verify this fact, measurements were made for the two-point correlation coefficients for 50 kinds of receiving point locations, except for within 1 m from the main noise source (computer). As a result, the measured correlation coefficient ranged from -0.15 to +0.15.

From the above result, it is seen that the noise in this field satisfies the condition for the proposed method i, i.e., the small two-point correlation coefficient. Despite the assumptions in the processing principle, however, the two-point correlation coefficient of speech decreases with the increase of the distance  $r$  between the speaker and the microphones. Consequently, for the three receiving conditions with different values of  $r$ , experiments were performed to verify the applicability of the proposed method.

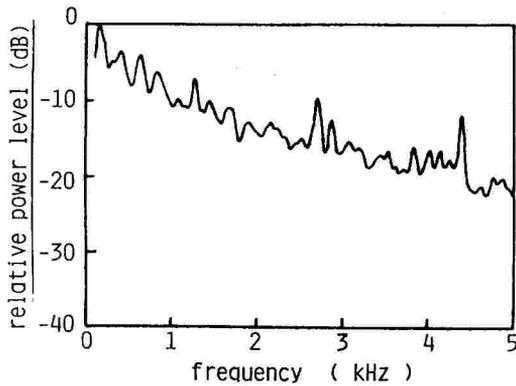


Fig. 7. Power spectrum of the room noise.

Figure 7 shows the power spectrum of the noise in the room. As is seen from this figure, the noise is a room noise with gradual spectrum changes with the slope of about  $-4.5$  dB/oct. However, it contains several periodic components.

#### 4.2 Conditions for sound receiving and articulation test

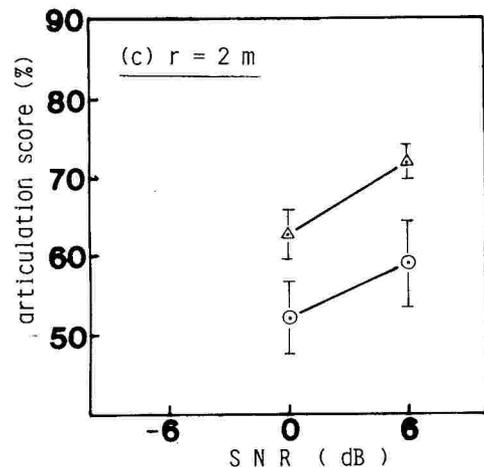
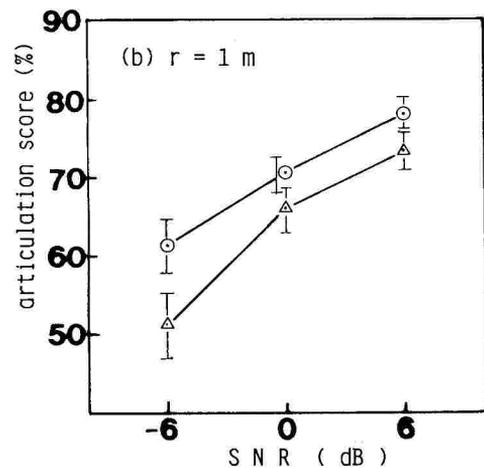
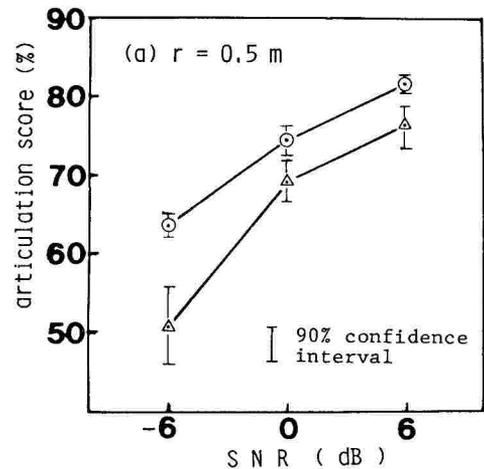
A loudspeaker was installed in the room described in the previous section, and 100 Japanese syllables were uttered. The sound was received by the two microphones shown in Fig. 5. The distance  $r$  between the loudspeaker and the microphones was set as 0.5, 1, and 2 m. Then the two-point correlation coefficients of speech were 0.95, 0.9, and 0.67, respectively. The two-point correlation coefficient of the noise at the microphone locations was approximately 0.1.

The signal-to-noise ratio is defined as the ratio of the average powers of speech and noise in voiced interval. By adjusting the output-level of the loudspeaker, the received sound was made at S/N ratio of +6, 0 and -6 dB. The listeners for the processed result were 4 females who are well-trained in the speech-quality test. The listening level was set at +3 dB OTR (orthotelephonic response) [10] at the average level of 100 syllables.

#### 4.3 Processing parameters

The received syllables were off-line processed for delay equalization (Fig. 3), after which the proposed method was applied. The parameters in the processing were as follows:

- (1) sampling frequency: 8 kHz;



△ : without processing  
 ○ : with processing

Fig. 8. Articulation score of the experiments.

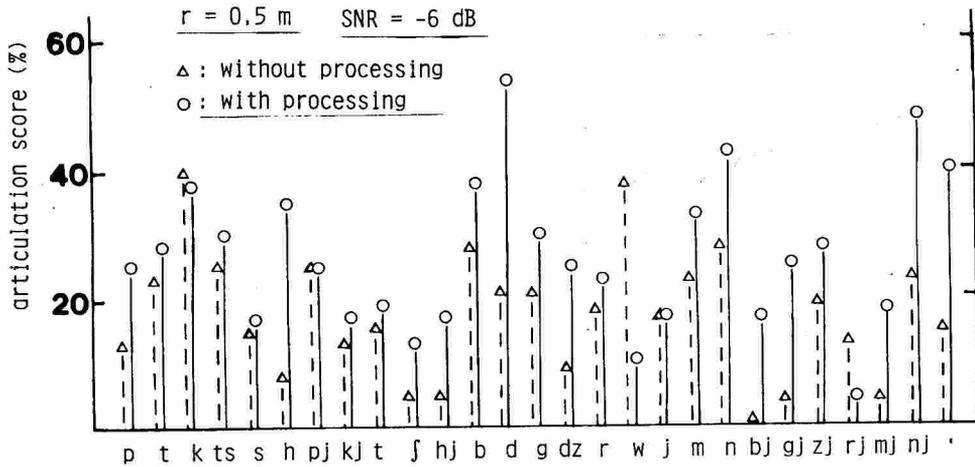


Fig. 9. Individual phonemes articulation score.

(2) Hamming window length for short-term Fourier transformation: 4 ms (32 points);

(3) averaging window  $w_R(m)$  for calculation of correlation coefficient:

Exponential window defined by

$$w_R(m) = \begin{cases} e^{-\frac{m}{L_c}} & m \geq 0 \\ 0 & m < 0 \end{cases} \quad (23)$$

where the time-constant denoted by  $L_c$  was set as 20 ms for S/N of 6 and 0 dB, and 80 ms for S/N of -6 dB; and

(4) Post-processing: Low-frequency components below 300 Hz were deleted by a high-pass filter.

#### 4.4 Experimental result

Figures 8(a) to (c) shows the results of articulation test, which are syllable articulation scores before and after processing. The ordinate in the figure is the syllable articulation score in percentage and the abscissa is the S/N ratio of the received speech. The triangle is the syllable articulation score without processing, and the circle is the syllable articulation score after noise suppression as given in this paper.

Figure 8(a) is the result of experiment when the distance  $r$  between the loudspeaker and the microphones is 0.5 m with the two-point correlation coefficient 0.95. It is seen from this figure that when the S/N ratio is +6 or 0 dB, the syllable articulation score can be increased by the proposed method by approximately 5%. When the S/N

ratio is as low as -6 dB, the syllable articulation score can be improved by approximately 13%.

Figure 8(b) is the result of experiment when the distance  $r$  is 1.0 m with the correlation coefficient of 0.9. Compared with the case of  $r = 0.5 \text{ m}$ , the effect of the proposed method is slightly decreased. Still, the syllable articulation score was improved by approximately 4% for S/N of +6 and 0 dB, and approximately 10% for S/N of -6 dB.

Figure 8(c) is the result when the distance  $r$  is 2 m with the two-point correlation coefficient 0.67. In this case, the syllable articulation score is decreased greatly by application of the proposed method. The two-point correlation coefficient of speech signal decreases with the increase of the distance between the loudspeaker and the microphones, and the correlation coefficient is decreased more remarkably for high-frequency components [12]. As a result, the high-frequency components of speech signal were suppressed by the processing and this may be the reason for the decrease of articulation score. However, the details will be investigated in the future.

The effectiveness of the proposed method was examined also from the viewpoint of phoneme. When  $r = 0.5$  or 1 m, no mishearing of vowels occurred for S/N of +6 and 0 dB both for before and after processing. When S/N is -6 dB, mishearing of vowels occurred by 16% before processing, which, however was decreased by the noise suppression processing to 4%. This indicates the effectiveness of the proposed method in improving the articulation of the vowel.

Figure 9 shows the individual phoneme hearing score of consonants. In the figure, triangle is the phoneme articulation score before processing, and the circle is that after noise suppression processing. The figure is the case where  $r = 0.5$  m and  $S/N = -6$  dB.

As a tendency observed in this figure, the proposed method is more effective for voiced bursts, such as b, d, g, bj and gj, and nasals such as m, n, mj and nj. The reason for this may be that the time-averaging in the determination of the correlation coefficient operates more effectively for the consonants with longer durations, and the noise suppression by weighting on the frequency-axis operates more effectively for the phoneme with the frequency spectrum with larger variations.

It was verified from the results of those experiments that the proposed method can improve the articulation if  $r = 1$  m, i.e., if the correlation coefficient of the noise is 0.1 or less and the two-point correlation coefficient of the speech is 0.9 or more. The correlation coefficient of speech is larger than 0.9 if the distance between the utterer and the microphone is 0.5 m or less under the room condition described in Sect. 4.1, as is seen from Fig. 6.

## 5. Conclusions

This paper proposed the new noise suppression signal processing method using the correlation coefficient of the two-point received signals. It is shown in the following that the articulation of speech can be improved by the processing.

(1) The noise suppression problem was discussed, assuming that there is no correlation among the desired signal and the noises received at the two different receiving points, also assuming that the reflective sound in the room has a negligible small level. As a result, it is shown that the noise suppression filter which is the optimum in the sense of the least-mean-square error can be obtained from the correlation coefficient of the frequency components of the two-point received signals.

(2) To apply the idea to the signal which can be regarded as stationary for a short interval, such as in the case of speech, the adaptive filtering based on the short-term correlation coefficient was proposed. Detailed procedures for noise suppression were presented.

(3) Finally, to verify the effectiveness of the proposed processing, articulation

tests were performed in the actual room sound field. Although the effectiveness depends on the distance between the utterer and the microphones as well as their locations, the method was verified to be effective if the noise field is diffused with two-point correlation coefficient 0.1 or less, and the two-point correlation coefficient of speech is 0.9 or more (those conditions are satisfied when the room volume is  $100 \text{ m}^3$ , reverberation time is 0.4 s, and the distance between the utterer and the microphones is 0.5 m or less), improving the syllable articulation score by 4 to 10%.

Acknowledgement. The authors acknowledge the suggestions made by Dr. Itakura, Head of Research Section 4, suggestions made in the articulation test by Dr. Kakei, and members of Comm. Stand. Group. They also acknowledge the suggestions made at the early stage of this work by Dr. Yoshikawa, former head of Research Section 2, and Dr. Koike, former head of Research Station 4.

## REFERENCES

1. M.M. Sondhi et al. Improving the Quality of a Noisy Speech Signal, *Bell Sys. Tech. J.*, 60, 8 (Oct. 1981).
2. M.W. Callahan. Acoustic Signal Processing Based on the Short-Time Spectrum, Ph.D. dissertation, Dep. Comput. Sci., Univ. Utah, Salt Lake City (Mar. 1976).
3. H. Nagabuchi. Suppression of random noise in speech based on the comb-filtering technique, *Tech. Rep. Speech, Acoust. Soc. Japan*, S79-48 (Nov. 1979).
4. K. Toshiya and J. Suzuki. Improvement in Signal-to-Noise Ratio by SPAC (speech processing system by use of autocorrelation function, *Trans. (A), I.E.C.E., Japan*, J61-A, 3, pp. 217-223 (Mar. 1978).
5. D.M. Chabries et al. Application of the LMS Adaptive Filter to Improve Speech Communication in the Presence of Noise, *Proceeding of ICASSP*, 1, pp. 148-150 (Mar. 1982).
6. J.S. Lim and A.V. Oppenheim. Enhancement and Bandwidth Compression of Noisy Speech, *Proc. I.E.E.E.*, 67, pp. 1586-1604 (Dec. 1979).
7. J.B. Allen, D.A. Berkley and J. Blauert. Multimicrophone Signal-Processing Technique to Remove Room Reverberation from Speech Signal, *J. Acoustic Soc. Am.*, 62, 4, pp. 912-915 (1977).
8. R.K. Cook et al. Measurement of Correlation Coefficients in Reverberant Sound Fields, *J. Acoust. Soc. Am.*, 27, 6, pp. 1072-1077 (Nov. 1955).
9. J.B. Allen. Short Term Spectral Analysis, Synthesis, and Modification by

Discrete Fourier Transform, I.E.E.E.  
Trans. Acoust., Speech and Signal Proc.,  
ASSP-25, pp. 235-238 (June 1977).

10. I.E.C.E., Japan, Speech and Hearing (1966).
11. M. Nakatsui, J. Suzuki and T. Takasugi. Subjective Ratio in SNR Improvement by SPAC, Proc. Conv. Acoust. Soc. Jap., 2-5-15 (Oct. 1978).
12. M. Tohyama and A. Suzuki. Interaural cross-correlation coefficients in reproduced sound fields, Tech. Rep. Arch., Acoust. Soc. Jap., AA81-27 (Sept. 1981).
13. M.B. Gardner. A Study of Talking Distance and Related Parameters in Hands-Free Telephony, Bell Sys. Tech. J., 39, 6 (Nov. 1960).