J. Acoust. Soc. Jpn. (E) 9, 3 (1988)

A method of speech recording in a noisy environment caused by vibration of panel

Akihiro Shima,* Chun-Duck Kim,** Akihiro Yuasa,*
Hiroshi Kanai,*** Masato Abe,* and Ken'iti Kido*

*Research Center for Applied Information Sciences, Tohoku University,
2-1-1, Katahira, Sendai, 980 Japan
**Pusan Fisheries University,
the Republic of Korea
***Education Center for Information Processing, Tohoku University,
Kawauchi, Sendai, 980 Japan

(Received 9 July 1987)

This paper summarizes the study of active cancellation of noise due to vibration of a panel at the output terminal of microphone. The vibration of panel, which radiates noise, is picked up by an acceleration pick-up. The noise output of the microphone due to vibration is estimated by computing the convolution between the output of acceleration pick-up and the transfer function from the pick-up to the microphone output. By subtracting the estimated noise output due to vibration, the noise in the microphone output is suppressed, and the necessary acoustic signal is left. The effectiveness and the limit of this method are experimentally shown. And the existence of zero points in the transfer function from the driving point to vibration pick up point due to vibration of panel is pointed out as the reason why the noise can not be completely suppressed when the vibration pick up point is far from the driving point.

PACS number: 43. 40. Dx, 43. 60. Gk, 43. 85. Ta

1. INTRODUCTION

This paper proposes a method of suppressing noise due to sound radiation by the vibration of panel at the microphone line.[1] The noise is estimated from the vibration of panel and the estimated noise is subtracted from the microphone output. This method is useful for picking up a speech signal in a noisy environment caused by panel vibration. First, a method of noise cancellation is described. Next, an expriment on noise cancellation using this method is outlined. The investigation is carried out on the relation between the length of time window for estimating the impulse response and the degree of cancellation, and on that between the vibration pick up point on the panel and the degree of cancellation.

We previously presented a paper[2,3] on the similar problem in which noise was radiated by a loudspeaker. The maximum decrease of 20 dB was achieved in the study, but the method was applicable only to noise radiated by a loudspeaker. The present method is more generally applicable, for example, it can be used in the case where the noise is radiated by wall or window.

2. PRINCIPLE

The noise is radiated by the vibration, x_p, of a panel, P. An acoustic signal such as speech is picked up by a microphone. Figure 1 shows the relation among these factors. Assuming the system is linear, the output signal of the microphone can be expressed as follows:

$$y_m = v + h_{pm} * x_p \,, \qquad (1)$$

where

x_p: the acceleration at point p on the vibrating panel

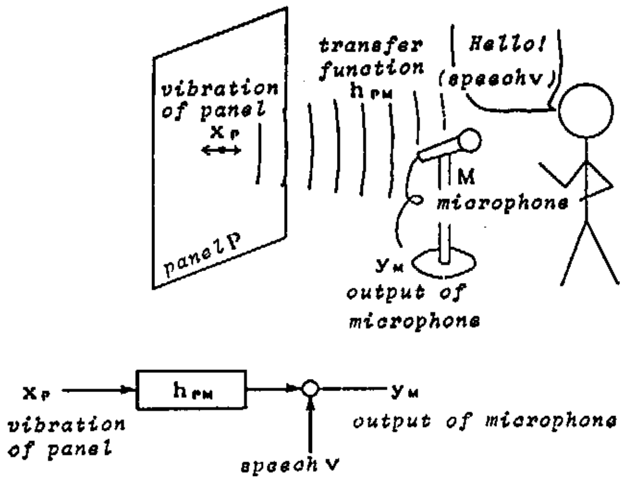

Fig. 1 Model of the problem.

v: the signal at the microphone output due to speech
y_m: the output signal of the microphone, M
h_{pm}: the impulse response of the microphone, M, to the vibration of panel at p
$*$: the convolution integral.

If the impulse response of the microphone, M, for the vibration of the panel can be estimated, the noise at the microphone output due to the vibration of panel is estimated by the computation of convolution between the vibration of panel and the impulse response. The necessary signal, v, can be extracted from the microphone output by subtracting the estimated noise as shown by the following equation:

$$\hat{v} = y_m - \hat{h}_{pm} * x_p , \tag{2}$$

where $\hat{v}$ and $\hat{h}_{pm}$ are the estimates of the speech and the impulse response, respectively. First, $\hat{h}_{pm}$ is estimated using the cross spectral method.[4] That is, the transfer function is estimated using the following equation:

$$\hat{H}_{pm} = \frac{\overline{X_p^* Y_m}}{\overline{X_p^* X_p}} , \tag{3}$$

where Y_m and X_p denote the discrete Fourier transform of y_m and x_p, X_p^* denotes the complex conjugate of X_p, and the overline denotes time averaging of sufficient number.

The estimated impulse response, $\hat{h}_{pm}$, is obtained by inverse Fourier transformation of the estimated transfer function, $\hat{H}_{pm}$. The convolution between $\hat{h}_{pm}$ and x_p represents the estimated noise component due to the vibration of panel at the microphone output. Therefore, the estimate of noiseless speech, $\hat{v}$, can be obtained by the two observable signals, x_p and y_m.

3. EXPERIMENT

Figure 2 shows the block diagram of the experimental setup. An experiment was carried out using one plywood panel with a thickness of 5 mm and another with a thickness of 18 mm. The plywood panel was driven by white noise using an electromagnetic shaker. The vibration of the panel was picked up by an acceleration pick-up fixed on the panel. The noise due to the vibration of the panel was picked up by a microphone which was set in front of the panel. The distance between the panel and the microphone was 1 m. These two signals were passed through low-pass filters with cut-off frequency of 2.3 kHz. The low passed signals were sampled at a period of 0.1 ms, A/D converted and stored in the memory of computer.

The length of time window for estimating the transfer function should be sufficiently longer than the reverberation time of panel. More accurate estimations of the impulse response can be achieved by a longer length of the time window. However, a longer time window causes a longer calculation time. Therefore, the relation between the accuracy of estimated impulse response and the length of time window was first investigated. That is, the two signals were cut out into many segments by Riesz window whose length was changed from 51.2 ms to 409.6 ms for the investigation. Each segment was Fourier transformed, and the power spectra and the cross spectra were calculated. Then, the transfer function was estimated using Eq. (3) after averaged. The transfer function was inversely Fourier transformed, and the first half section was used as the impulse

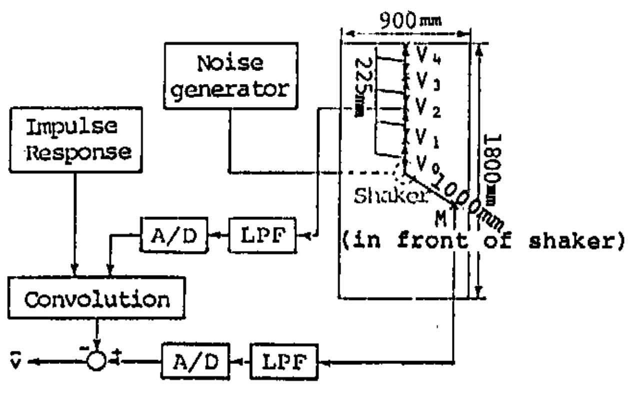

Fig. 2 Experimental system.

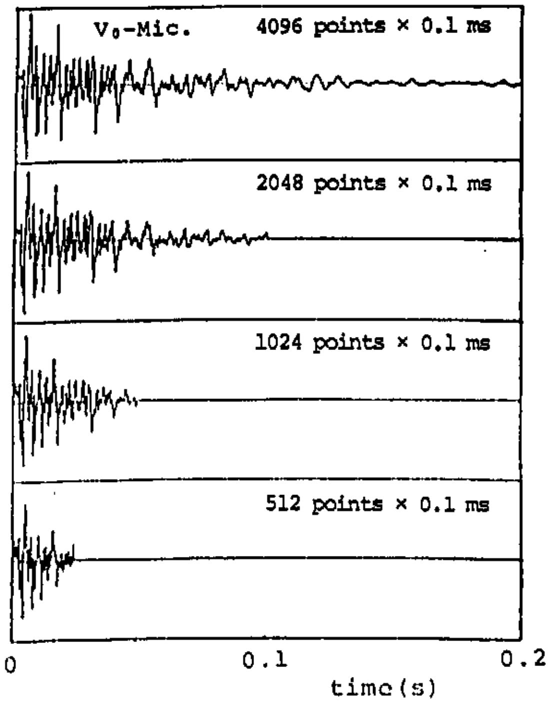

Fig. 3 Estimated impulse responses for the panel with a thickness of 18 mm using various length of time window.

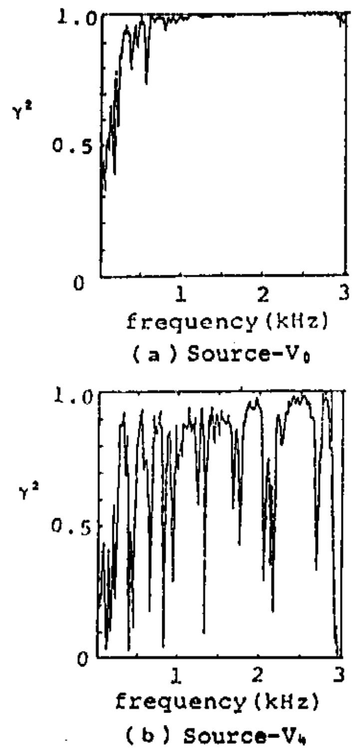

Fig. 4 Coherence function between the vibration driving source and the output of vibration pick-up (thickness of panel: 5 mm).

response.

The Hanning window is generally used to minimize the edge effect of time window for the estimation of transfer function. However, the Hanning window offers poor estimation accuracy when it is used to estimate an impulse response whose length is nearly the same as that of the time window. It has already been shown that it is better to use a time window like the Riesz window or the half cosine window,[5] whose weighting factors rapidly decrease at both ends of the window.

Figure 3 shows the impulse response estimated by the above described method where a panel with a thickness of 18 mm was used. In this case, even for 2,048 points (204.8 ms) the impulse response was not sufficiently attenuated. This means that a much longer time window is necessary to achieve an accurate estimation. The estimated impulse response is convolved with the output signal of the vibration pick-up, and the convolution is subtracted from the microphone output. If the system is linear and the estimation of impulse response is accurate, the noise due to the vibration of panel should be well suppressed.

Figure 4 shows the coherence function between the vibration driving source and the outputs of vibration pick-ups at V_0 and V_4, where the panel thickness is 5 mm and the average time is 128. The value of the coherence function is nearly one when the pick-up point is at the driving point, V_0, however, the value is small at the point V_4 which is far from the driving point. This means that uncorrelated components are large at the edge of panel. If the value of the coherence function is one, it can be assumed that the transfer function is accurately estimated, and a considerable noise suppression can be expected. However, the value of the coherence function drops at many frequencies, especially in the case of V_4 where the vibration pick-up point is far from the driving point. The followings are considered to be the reason of the drops: 1) the external noise, 2) the non-linear vibrations of the panel, 3) existence of zero points in the transfer function from the driving point to the pick-up point.

Figures 5 and 6 show the coherence functions between vibration of the panel and the output of microphone, where the thickness of the panel is 5 mm and 18 mm, respectively. The value of the coherence

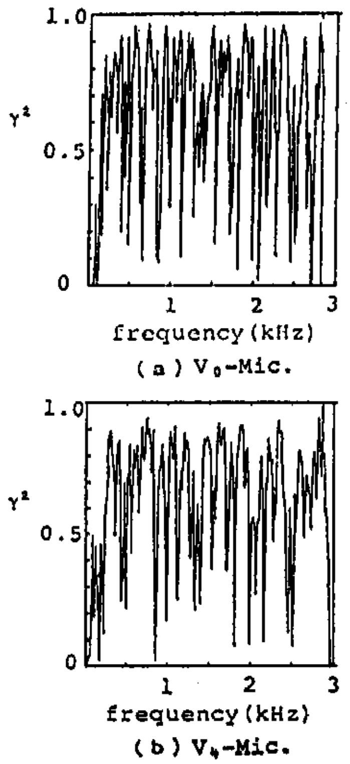

Fig. 5 Examples of coherence function between the vibration of panel and the output of microphone (thickness of panel: 5 mm).

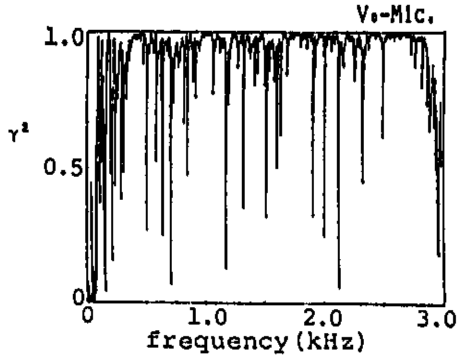

Fig. 6 An example of coherence function between the vibration of panel and the output of microphone (thickness of panel: 18 mm).

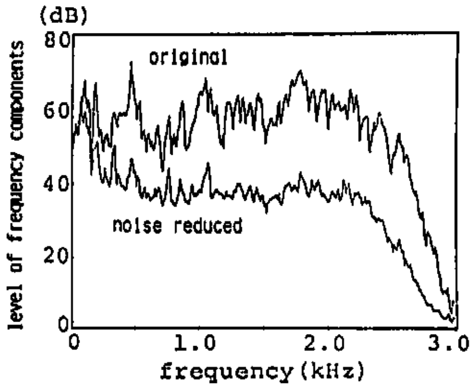

Fig. 7 Changes in the noise spectrum resulting from the cancellation (thickness of panel: 18 mm).

function obtained with the 18 mm thickness panel is higher than that obtained with the 5 mm thickness panel. The reason is described in the next section. Figure 7 shows an example of noise suppression by the frequency spectra of noise in the microphone output, where the thickness of panel is 18 mm and the pick up point is V_0. In this case, the coherency is sufficiently high (as shown in Fig. 6), so that the noise can be well suppressed over most of the frequency range.

Figure 8 shows the relation between the length of time window used to estimate the impulse response and the degree of noise suppression for the cases of 5 mm thickness panel and 18 mm thickness panel. It is also clear from Fig. 3 that the accuracy in the estimation of impulse response varies according to the length of time window. But in Fig. 8, the effect of the length of time window can be seen only in the case where the vibration is picked up at V_0 and the thickness of the panel is 18 mm. In the other cases, the degree of noise suppression can hardly be changed with the increase in the length of time window. It seems that there are some other factors than the length of time window to make fruitless the noise suppression by this method.

Figure 9 shows the relation between the degree of noise suppression and the pick up point which is moved from V_0 to V_4 in Fig. 2. In all cases, the noise suppression is less when the vibration pick-up is set far from the driving point, V_0. In the case of 18 mm thickness, the noise suppression is better than that in the case of 5 mm thickness. However, the change in the noise suppression by the vibration pick-up point is greater in the case of 18 mm thickness than that in the case of 5 mm thickness.

4. EFFECT OF ZERO POINTS ON THE ESTIMATION OF OUTPUT NOISE

Figure 10 shows a model of system including zero points in the transfer function from the driving point to the vibration pick up point. Where X_0, X_1, Y and

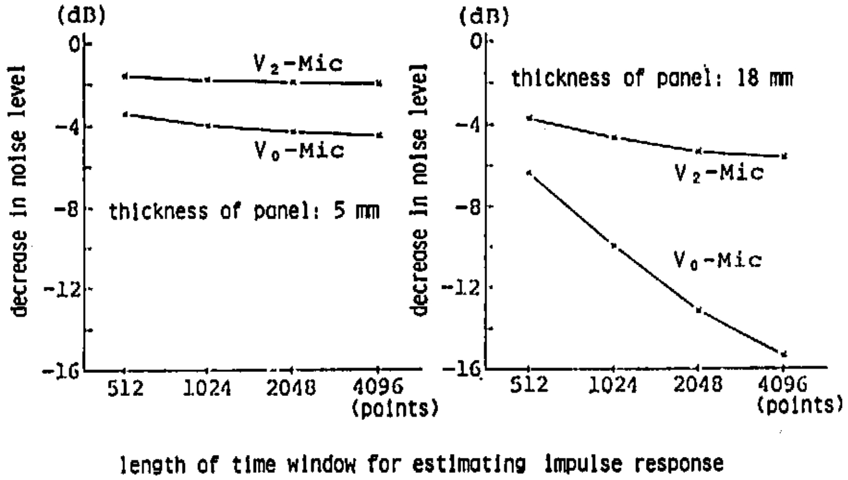

Fig. 8 Relation between the length of time window and the degree of noise suppression.

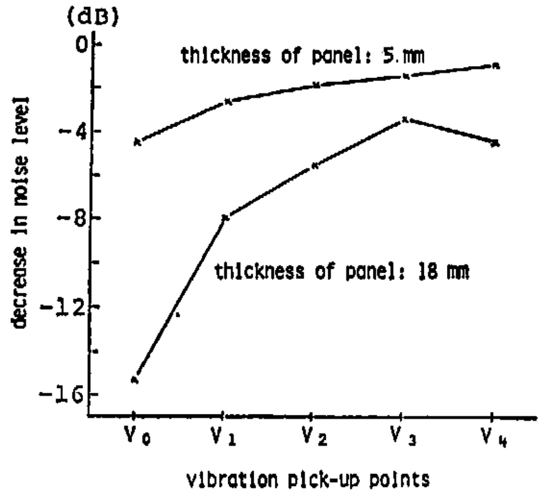

Fig. 9 Relation between the vibration pick-up point and the degree of noise suppression.

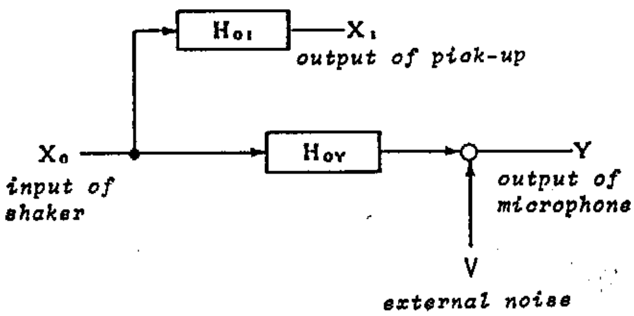

Fig. 10 Model of the system where X_1 is observed instead of source signal X_0 (X_1 can include zero points because H_{01} is transfer function through panel vibration).

V are the DFT of the input of shaker, the output of vibration pick-up on the vibrating panel, the output

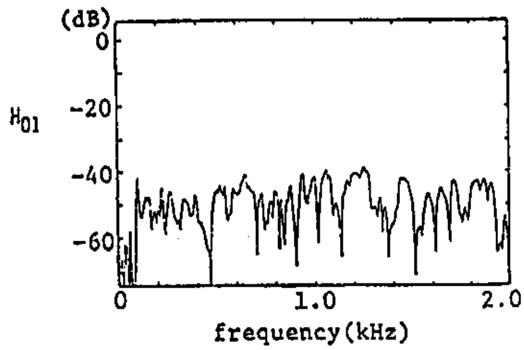

Fig. 11 An example of experimentally obtained transfer function H_{01} including zero points.

signal of the microphone and the external signal, respectively. H_{01} and H_{0Y} are the transfer functions from shaker to pick-up and that from shaker to microphone, respectively.

It is first assumed that H_{01} has zero points as there are many nodal points on the panel and the nodal points coincide often with the vibration pick up point. The output signal of the microphone is expressed as follows:

$$Y = H_{0Y}X_0 + V = \frac{H_{0Y}}{H_{01}}X_1 + V. \qquad (4)$$

At the frequencies where H_{01} and X_1 are zero, Y becomes unstable, and (H_{0Y}/H_{01}) cannot be estimate. That is, it is impossible to cancel the noise using the estimated impulse response at those frequencies. Under certain circumstances, noise may even increase in the microphone output on the contrary. Figure 11 shows the example of transfer function from the vibration driving source to the output of vibration pick-up. There are many dips of more than 20 dB at some frequencies.

Theoretically it may be possible to estimate the output Y unless H_{01} has complete zero points in the spectrum. However, in the practical case, it is difficult to estimate the output accurately because of the presence of external noise, non-linear vibrations and the effect of the time window as explained below.

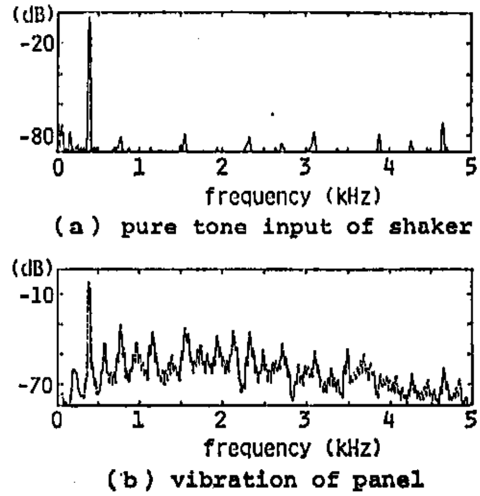

Fig. 12 Examples of the spectra of panel vibration where the effect of non-linear vibration is observed.

(1) The effect of non-linear vibrations

Figure 12 shows the power spectrum of signal which drives the shaker and that of the vibration signal of panel in the case where pure tone is the input of shaker. The high frequency components due to the non-linear vibrations are more than 30 dB lower than the input power of sinusoidal wave. Even if the components uncorrelated to input signal is so small as shown in the figure, the coherence function will decrease in the value at the frequencies of zero points in the transfer function H_{01}.

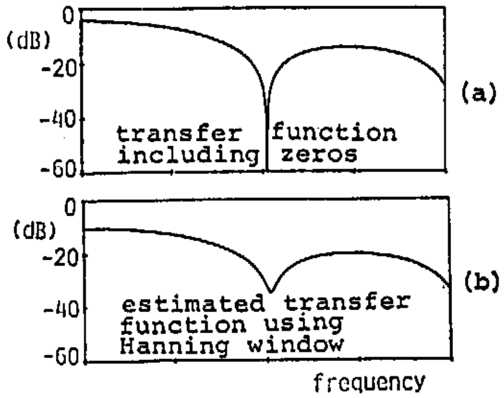

Fig. 13 Example of estimation error at zero point in transfer function due to the effect of time window.

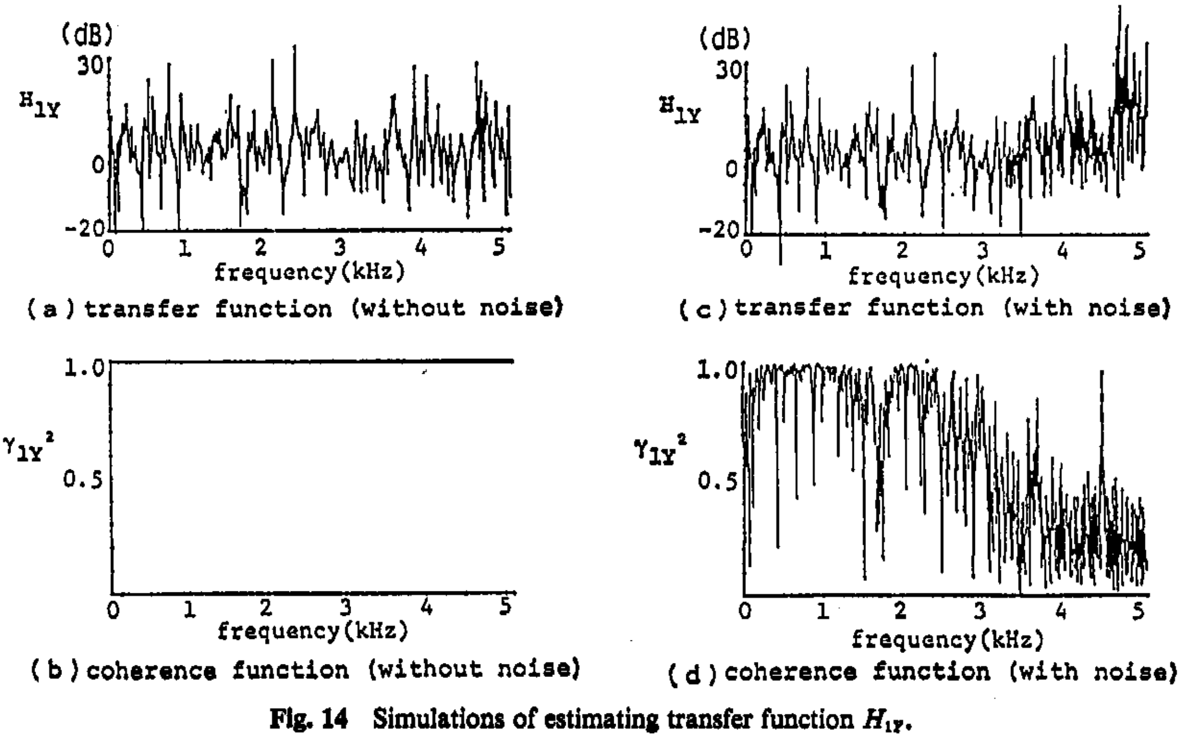

Fig. 14 Simulations of estimating transfer function H_{1Y}.

(2) The effect of time window

Figure 13 (a) shows the spectrum of signal having a zero point in the spectrum, and Fig. 13 (b) shows the spectrum of the same signal cut out by Hanning window. The dip shown in Fig. 13 (a) becomes shallower and broader. The sharp dips in transfer function is often burried in the process of cross spectral estimation using sequence of short duration. That is, the transfer function cannot be accurately estimated by use of the time window of short duration.

(3) The effect of external noise

The effect is investigated by computer simulation. The experimentally obtained transfer function H_{01} from the vibration driving source to the output of the vibration pick-up is used as the transfer function for the simulation. Using computer generated white noise as the input X_0 of the shaker, the output of pick-up X_1 and the output of microphone Y are calculated. In the case where there is no external noise, the transfer function H_{1Y} and the coherence function γ_{1Y}^2 between the vibration of panel and the output of microphone are correctly estimated as shown in Figs. 14 (a) and 14 (b), respectively. For the case where the white noise of -20 dB is mixed to the output signal, the estimated H_{1Y} and γ_{1Y}^2 are shown in Figs. 14 (c) and 14 (d), respectively. Without external noise, the coherence function is one at all frequencies and the estimated transfer function H_{1Y} is reliable. However, in the case where the noise is mixed, coherency decreases at the frequencies of zero points and an accurate estimation can not be expected at those frequencies.

$\hat{Y}$ is estimated using the cross spectrum technique. The normalized mean square value of the difference between $\hat{Y}$ and Y is -96.6 dB in case of no external noise, and -18.9 dB in case of -20 dB external noise. The results of the simulation show that it is difficult to estimate the transfer function at frequencies of zero points even in the case of slight external noise.

5. CONCLUSION

The principle of suppression of noise at the microphone output terminal is described with some experimental results where the noise is radiated by vibrating panel and is picked up by the microphone.

In the proposed method, the transfer function from the vibration of panel to the microphone output is first estimated. Then, the noise in the microphone output is estimated by computing the convolution between the vibration signal and the estimated impulse response. And the estimated noise is subtracted from the microphone output. The results of our experiments show that the noise is sufficiently reduced in case of the panel of 18 mm thickness when the pick-up point is close to the driving point. It is also found that a longer time window for the estimation of the impulse response gives better results. However, only poor noise suppression is obtained when the panel is thin or the vibration pick-up point is far from the driving point of the panel. Experiments show that the main reasons for this poor noise suppression is the effect of zero points in the transfer function through the panel.

How to avoid the effect of the zero points to achieve better results will be presented in the near future.

REFERENCES

1) C. D. Kim, S. Morikawa, M. Abe, and K. Kido, "Active suppression of noise in microphone line using vibration of noise source," Inter-Noise 84 Proc., 469–472 (1984).
2) C. D. Kim, M. Abe, and K. Kido, "Cancellation of signal picked up in a room by estimated signal," Inter-Noise 83 Proc., 423–426 (1983).
3) C. D. Kim, M. Abe, and K. Kido, "Cancellation of signal picked-up in a room by estimated signal," 11th ICA Proc., 301–304 (1983).
4) G. M. Jenkins and D. G. Watts, *Spectral Analysis and Applications* (Holden-Day, San Francisco, 1968).
5) M. Abe, C. D. Kim, and K. Kido, "Investigation of the effect of a time window on the accuracy of an estimated impulse response," J. Acoust. Soc. Jpn. (E), 7, 269–277 (1986).