Information about the distance to the target is very important in a number of engineering fields. We previously proposed an acoustic distance measurement (ADM) method based on the interference between the transmitted and reflected waves, which can be used for distance measurements over a wide range, from short range to long range. The proposed method assumes that the magnitude of the frequency response of the path from the loudspeaker to the microphone (i.e., the frequency response of the measurement system) is flat. However, in actual measurements, this is difficult to achieve. The lack of flatness of the magnitude of the frequency of the measurement system affects the observed sound, primarily in the low-frequency region; this results in incorrect estimations of the distance, especially at close range. In order to remove the effect of the measurement system, our previous methods required two measurements, i.e., with and without targets. Thus, we proposed an ADM method that is based on the phase interference obtained by using the cross-spectral method with stereo microphones. This method does not require that the magnitude of the frequency of the measurement system is flat. The conventional cross-spectral method requires the placement near a loudspeaker of both a measuring remote microphone and a reference microphone. In contrast, the proposed method requires stereo microphones to be placed near the sound source. However, the performance of the proposed method was degraded by a noisy environment. In the present paper, the ADM method, which is based on the phase interference obtained using the cross-spectral method with stereo microphones, is expanded to a method that is robust to noisy environments. In order to measure distances in an environment with a low signal-to-noise ratio (low SNR), we introduce a synchronous addition to the proposed method. Finally, we confirmed the validity and effectiveness of the newly proposed method in real environments.
1. Introduction

Information about the distance to the target is very important in a number of engineering fields. In particular, distances must be known for the practical use of hands-free speech interfaces and nursing-care robots. A number of distance measurement methods, which use the time delay of a reflected wave measured with reference to the transmitted wave, have been proposed [1, 2]. However, these methods cannot be used to measure short distances because the transmitted wave will not be sufficiently attenuated when the reflected wave is received and will suppress the reflected wave [3, 4, 5]. This is caused by the transient response characteristics of the equipment in real environments. Thus, even if an impulse response can be measured, the distance cannot be measured when the distance is short. For this reason, we previously proposed an acoustic distance measurement (ADM) method based on the interference between the transmitted and the reflected waves [6, 7, 8], which can be used for measurements over a wide range of distances, from short range to long range.

The method proposed in this paper assumes a priori information about the transmitted sound and the flatness of the magnitude of the frequency response of the path from the loudspeaker to the microphone (i.e., the frequency response of the measurement system). However, in actual measurements, even if the frequency characteristics of the transmitted sound can be known in advance (for example, if the transmitted sound is emitted artificially), it is very difficult to attain sufficient flatness. The lack of flatness in the magnitude of the frequency of the measurement system affects the observed sound, primarily in the low-frequency region. This results in incorrect estimations of the distance, especially at close range. In order to remove the effect of the measurement system, our previous methods required two measurements, i.e., with and without targets [6, 7]. The cross-spectral method has been proposed as a method for measuring the acoustic transfer function [9]. This method can measure the transfer function between two microphones by using reference and measurement microphones. Thus, we also proposed an ADM method based on the phase interference obtained using the cross-spectral method with stereo microphones [10], and this method did not require the magnitude of the frequency of the measurement system to be flat. The conventional cross-spectral method requires the placement near the loudspeaker of a measuring remote microphone and a reference microphone. In contrast, the proposed method requires stereo microphones to be placed near the sound source. However, the performance of the proposed method was degraded in a noisy environment.

In the present paper, the ADM method based on the phase interference obtained using the cross-spectral method with stereo microphones is expanded to a method that is robust to noisy environments. In order to measure the distance in an environment that has a low signal-to-noise ratio (low SNR) environment, we introduce a synchronous addition to the proposed method. Finally, we confirmed the effectiveness of the newly proposed method by performing experiments in real environments.

2. Theoretical Consideration

2.1 Acoustic distance-measurement method based on interference and using the cross-spectral method [10]

In this section, we summarize the principle of the acoustic distance-measurement method based on phase interference obtained by using the cross-spectral method with stereo microphones.

Figure 1 shows the relative positions of the two microphones, the sound source (loudspeaker), and the target. As shown in Fig. 1, microphones 1 and 2 are set, respectively, at the positions \( x = x_1 \) [m] and \( x_2 \) [m], where the horizontal axis is the \( x \)-axis, the origin is set at the position of microphone 1 (\( x_1 = 0 \) m), and the target is set at \( x = d \) [m].

Let the transmitted wave \( v_T(t, x_k) \), which expresses the sound pressure, be a function of position...
$x_k$ [m] ($k = 1, 2$) and time $t$ [s], as follows:

$$v_T(t, x_k) = \int_{f_1}^{f_N} A(f) e^{j(2\pi ft - \frac{2\pi ft x_k}{c})} df,$$

(1)

where $f$ [Hz] is the frequency, $f_1$ [Hz] and $f_N$ [Hz] correspond to the lowest and highest frequencies, respectively, $A(f)$ is the spectrum of the transmitted wave, and $c$ [m/s] is the speed of sound.

For simplicity, assuming that the transmission wave is reflected by one target, the reflected wave is represented as follows:

$$v_R(t, x_k) = \int_{f_1}^{f_N} A(f) \gamma(f) e^{j(2\pi ft - \frac{2\pi ft}{c}(2d - x_k) + \phi(f))} df,$$

(2)

where $\gamma(f) e^{j\phi(f)}$ is the reflection coefficient of the target and $d$ is the distance [m] to be estimated (between the target and microphone 1).

As shown in Fig. 1, the measurement system with an impulse response $h(t)$ (consisting of the impulse response $h_L(t)$ of the audio playback system and the impulse responses $h_{M1}(t)$ and $h_{M2}(t)$ of the audio recording systems) affects the sound observations. Here, $h_{M1}(t)$ can be assumed to be approximated by $h_{M2}(t)$ (i.e., $h_M(t) = h_{M1}(t) \approx h_{M2}(t)$). For a single target, the composite wave (the composition of all transmitted and reflected waves) at $x_1 (= 0$ m) and $x_2$ m, is formulated as

$$v_C(t, x_k) \approx h(t) * \{v_T(t, x_k) + v_R(t, x_k)\},$$

(3)

$$h(t) = h_L(t) * h_{M1}(t),$$

(4)

where $*$ is the convolution operator. By applying the Fourier transform to $v_C(t, x_k)$, the Fourier spectrum $V_C(f, x_k)$ can be easily obtained as follows:

$$V_C(f, x_k) = \int_{-\infty}^{\infty} v_C(t, x_k) e^{-j2\pi ft} dt$$

$$= A(f) H(f) e^{-j\frac{2\pi ft x_k}{c}} + A(f) H(f) \gamma(f) e^{-j\frac{2\pi ft}{c}(2d - x_k) - \phi(f)},$$

(5)

where $H(f)$ is the frequency response of the measurement system.

In order to detect the interference, by regarding $v_C(t, x_1)$ and $v_C(t, x_2)$ as the input and output signals, respectively, in the cross-spectral method, the cross-spectrum is obtained as follows:

$$C(f, x_1, x_2) = \frac{V_C^*(f, x_1)V_C(f, x_2)}{V_C^*(f, x_1)V_C(f, x_1)},$$

(6)

where $V_C^*(f, x_1)$ represents the complex conjugate of $V_C(f, x_1).$
When the observation point is located near a sound source, we can assume that \( \gamma \ll 1 \). Thus, \( C(f, x_1, x_2) \) is approximated as follows:

\[
C(f, x_1, x_2) \approx \frac{e^{-jD(f)} + \gamma(f)e^{-j\beta(f)} + \gamma(f)e^{j\beta(f)}}{1 + \gamma(f)e^{-j\alpha(f)} + \gamma(f)e^{j\alpha(f)}},
\]

(7)

\[
D(f) = \frac{2\pi f}{c} x_2,
\]

(8)

\[
\alpha(f) = \frac{2\pi f}{c} 2d - \phi(f),
\]

(9)

\[
\beta(f) = \frac{2\pi f}{c} (2d - x_2) - \phi(f).
\]

(10)

Furthermore, when \( \gamma(f) \) can be assumed to be constant (\( \gamma(f) = \gamma \)) and \( \gamma \ll 1 \), the power of \( C(f, x_1, x_2) \) is approximated as follows:

\[
p(f, x_1, x_2) = |C(f, x_1, x_2)|^2 \\
\approx 1 + 2\gamma \left\{ -\cos \left( \frac{4\pi f}{c} d - \phi(f) \right) + \cos \left( \frac{4\pi f}{c} (d - x_2) - \phi(f) \right) \right\},
\]

(11)

where the two cosine terms indicate the phase interference. Thus, \( p(f, x_1, x_2) \) is a periodic function, and this period is inversely proportional to the distances \( d \) [m] and \( d - x_2 \) [m] between the observation points and the target. Thus, we can estimate the distances \( d \) [m] and \( d - x_2 \) [m] by applying the Fourier transform to \( p(f, x_1, x_2) \). Specifically, in order to extract only the periodic function, we subtract the average \( p(f, x_1, x_2) \) of \( p(f, x_1, x_2) \) as follows:

\[
\Delta p(f, x_1, x_2) = p(f, x_1, x_2) - \overline{p(f, x_1, x_2)}.
\]

(12)

\( \Delta p(f, x_1, x_2) \) is referred to as the delta power of the cross-spectrum.

Consequently, the distances between the observation points and the target can be determined by again applying the Fourier transform to \( \Delta p(f, x_1, x_2) \). Namely, in the Fourier transform formula

\[
(G(f) = \int_{-\infty}^{\infty} g(t)e^{-j2\pi ft} dt),
\]

by replacing \( f \) with \( 2\pi/c \), \( t \) with \( f \), and \( g(t) \) with \( \Delta p(f, x_1, x_2) \), \( P(x) \) can be obtained by the following formula:

\[
P(x) = \int_{f_1}^{f_N} \Delta p(f, x_1, x_2)e^{-j2\pi \frac{x}{c} f} df,
\]

(13)

where this transform differs from the cepstrum[11] in that this transform is the Fourier transform and not its inverse. The peak positions of the range spectrum \( |P(x)| \) correspond to the distances \( d \) [m] and \( d - x_2 \) [m], which are to be estimated. When the difference \( |d - (d - x_2)| = |x_2| \) between two estimated values (i.e., the microphone interval) is sufficiently small, \( d \) [m] and \( d - x_2 \) [m] are assumed to be equal. Therefore, the two peaks of \( |P(x)| \) overlap, and only one peak appears at the position of \( d \) [m].

An additional result is that the minimum measurable distance \( d_{\text{min}} \) is defined by the frequency bandwidth \( f_W (= f_N - f_1)[6, 7] \). Namely, the period of \( C(f) \) must be shorter than \( f_W \) in order to precisely determine the peaks of \( |P(x)| \). Thus,

\[
d_{\text{min}} = \frac{c}{2f_W}.
\]

(14)

2.2 Noise reduction using synchronous addition

In an actual measurement, an observation sound \( v_c(t, x_k) \) at the \( k \)-th microphone position \( x_k \) is inevitably contaminated by a background noise \( w_k(t) \) as follows:

\[
v'_c(t, x_k) = v_c(t, x_k) + w_k(t) \quad (k = 1, 2).
\]

(15)
The background noise degrades the performance of the ADM method, because it does not use the observation sound \( v_C(t, x_k) \) itself but instead uses the phase interference terms in Eq.(11). The phase interference terms are multiplied by the magnitude of the reflection coefficient and take small values, so that the background noise significantly affects the performance. Although it would be better to define the SNR as the ratio of the reflected sound \( v_R(t, x_k) \) to the noise \( w_k(t) \) instead of the ratio of the observation sound \( v_C(t, x_k) \) to the noise \( w_k(t) \), the reflection coefficient is unknown.

When \( w_k(t) \) is negligible compared with the reflected wave \( v_R(t, x_k) \) (i.e., the SNR is high), the range spectrum \(|P(x)|\) clearly has a peak at the desired distance. In the low-SNR case, the true peak of \(|P(x)|\) is buried in spurious peaks that are caused by the background noise. Thus, it is essential to reduce the effect of background noise in order to measure the distance between the microphone and the target.

![Figure 2. Illustration of the time-shifted frame for the noisy observation \( v'_C(t, x_k) \).](image)

Let \( v'_C(t, x_k; s) \) be the \( s \)-th frame, shifted in the time domain, for the noisy observation, as shown in Fig. 2. In order to reduce the noise, we introduce synchronous addition in the frequency domain. The cross-spectrum between the noisy observations \( v'_C(t, x_1; s) \) and \( v'_C(t, x_2; s) \) is obtained as follows:

\[
C'(f, x_1, x_2; s) = \frac{V_{C'}(f, x_1; s)V_{C'}(f, x_2; s)}{V_{C'}(f, x_1; s)V_{C'}(f, x_1; s)},
\]

\[
= \left\{ V_C(f, x_1; s) + W_1(f; s) \right\}^* \left\{ V_C(f, x_2; s) + W_2(f; s) \right\} \left\{ V_C(f, x_1; s) + W_1(f; s) \right\}^* \left\{ V_C(f, x_1; s) + W_1(f; s) \right\},
\]

where \( W_k(f; s) \) is the Fourier spectrum of the background noise at the \( k \)-th microphone in the \( s \)-th frame.

When synchronous addition is applied in the frequency region to both the denominator and the numerator of the cross-spectrum \( C'(f, x_1, x_2; s) \), which is based on the deterministic functions \( V_C(f, x_1; s) \) and \( V_C(f, x_2; s) \), the terms based on the background noises \( W_1(f; s) \) and \( W_2(f; s) \) become small since they are inversely proportional to the number of frames \( S \) of the synchronous addition. It is expected that the synchronous addition will reduce the effect of the noise on the accuracy of the estimation of the distance.

### 3. Verification of the Proposed ADM Method

#### 3.1 Evaluation experiment in a real environment

In order to examine the effect of the observation noise on the current proposed model and to consider a way to counter the background noise, we carried out evaluation experiments in real
environments under various SNR conditions.

3.1.1 Experimental conditions

The experimental conditions and the experimental apparatus are shown, respectively, in Tables 1 and 2. In Table 1, the SNR = $\infty$ dB, which indicates a clean environment. We adopted a Gaussian white noise for the background noise. Microphone 1 is set at the origin, and the distances between the loudspeaker and microphone 1, between microphone 1 and the target, and between microphone 1 and microphone 2 are 0.5 m, 1 m, and 0.006 m, respectively. We carried out the experiment in a room (whose volume is 6.0 m $\times$ 7.0 m $\times$ 3.0 m) without obstacles.

Table 1. Experimental conditions.

<table>
<thead>
<tr>
<th>Transmitted wave source</th>
<th>Random phase</th>
</tr>
</thead>
<tbody>
<tr>
<td>Noise source</td>
<td>Gaussian white noise</td>
</tr>
<tr>
<td>SNR</td>
<td>-5, 0, 5, 10, $\infty$ dB</td>
</tr>
<tr>
<td>Sampling frequency</td>
<td>44.1 kHz</td>
</tr>
<tr>
<td>Quantization</td>
<td>16 bit</td>
</tr>
<tr>
<td>Measurement time</td>
<td>47 ms</td>
</tr>
<tr>
<td>Frame length</td>
<td>2048 samples</td>
</tr>
<tr>
<td>Frame shift</td>
<td>1 sample</td>
</tr>
<tr>
<td>Frequency bandwidth</td>
<td>5.5 kHz (2.1 kHz $\sim$ 7.6 kHz)</td>
</tr>
<tr>
<td>Minimum measurable distance</td>
<td>0.03 m</td>
</tr>
<tr>
<td>Sound Speed</td>
<td>340 m/s</td>
</tr>
<tr>
<td>Reverberation time</td>
<td>0.3 s</td>
</tr>
</tbody>
</table>

Table 2. Experimental apparatus.

<table>
<thead>
<tr>
<th>Target</th>
<th>Plywood square</th>
</tr>
</thead>
<tbody>
<tr>
<td>Audio interface</td>
<td>ROLAND, UA-55</td>
</tr>
<tr>
<td>Loudspeaker</td>
<td>YAMAHA, MSP5 STUDIO</td>
</tr>
<tr>
<td>Microphone</td>
<td>AUDIO-TECHNICA, AT9904</td>
</tr>
<tr>
<td>Microphone amplifier</td>
<td>PAVEC, MA-2016C</td>
</tr>
</tbody>
</table>

Figure 3. Experimental environment.

3.1.2 Evaluation index

We used the distance detection ratio (DDR) as the evaluation for the detection rate of the distance, as estimated from the distance spectrum. In this experiment, the distance measurement was done 100 times for each SNR condition, and DDR [%] is the ratio of the times that the peak is detected within the allowable error ($\pm d_{\text{min}}$ = $\pm$0.03 m) when the target is located at $d = 1$ m) to the total number of peaks. To calculate the DDR, the denominator and the numerator of the cross-spectrum $C''(f, x_1, x_2, s)$ are synchronously added in the frequency region.
### 3.1.3 Experimental results by the conventional method

The DDRs for the previously proposed ADM method are shown in Fig. 4 for various SNRs. It was found that even the conventional method is robust to noise in a high-SNR environment (e.g., \( \text{SNR} = \infty \) or greater than 10 dB). However, since the DDR is decreased around \( \text{SNR} = 7 \) dB, in that range, it is difficult to measure the distance due to the effects of the low-SNR environment, and the background noise degrades the performance of the previously proposed ADM method.

![Distance detection rate](image)

**Figure 4.** Distance detection rate.

### 3.2 Noise suppression

By applying a Fourier transform to the observations \( v'_C(t, x_1, 1) \) and \( v'_C(t, x_2, 1) \) in the first frame, the cross-spectrum for that frame can be obtained as in Eq.(16). Shifting the frames of the observations by \((s - 1)\) \( (v'_C(t, x_1, 1) \) and \( v'_C(t, x_2, 1) \) to \( v'_C(t, x_1, s) \) and \( v'_C(t, x_2, s) \)) and applying a Fourier transform to \( v'_C(t, x_1, s) \) and \( v'_C(t, x_2, s) \) yields the cross-spectrum for the \( s \)-th frame. In order to suppress the noise in a low-SNR environment, we applied synchronous addition to both the denominator and the numerator of the cross-spectrum in Eq.(16). Figure 5 shows the DDRs for various SNRs, as obtained by applying synchronous addition, where the number of frames of synchronous addition is 1, 3, 5, and 10. It is obvious that in a low-SNR environment, we can obtain a high DDR by applying synchronous addition to both the denominator and the numerator of the cross-spectrum.

![Distance detection rate (synchronous addition)](image)

**Figure 5.** Distance detection rate (synchronous addition).
4. Conclusion

In the present paper, we have expanded the previously proposed ADM method, which is based on the phase interference obtained by using the cross-spectral method with stereo microphones, to a method that is robust to noisy environments. In order to measure the distance in an environment in which the signal-to-noise (SNR) is low, we have introduced synchronous addition to the proposed method. By performing experiments in real environments, we have confirmed that the performance was improved. In our future work, we intend to reduce the effect of the background noise by using only a single frame of observations, instead of the averaging operation over multiple observations, as is used in the current method.

REFERENCES


