Smartphone Localization using Active–Passive Acoustic Sensing

Hiroaki Murakami, Masanari Nakamura, Shoma Yamasaki
Hokkaido University
Sapporo, 060-0814 Japan
Email: {hmurakami@ist, nmasanari@frontier, yshoma@ist}.hokudai.ac.jp

Hiromichi Hashizume
National Institute of Informatics
Tokyo, 101-8430 Japan
Email: has@nii.ac.jp

Masanori Sugimoto
Hokkaido University
Sapporo, 060-0814 Japan
Email: sugi@ist.hokudai.ac.jp

***Abstract*—In this paper, we describe a novel position-recognition method that uses passive acoustic signals from two previously installed speakers (passive acoustic sensing) and active acoustic signals from a smartphone's loudspeakers (active acoustic sensing). In passive acoustic sensing, a locus of positions for the smartphone can be calculated from the measured time difference of arrival from the two installed speakers. In active acoustic sensing, a chirp signal is transmitted from the speakers of the smartphone, and the distance to the side wall is measured from the propagation time of arrival at its microphone. We can obtain the smartphone position from our proposed model equations by combining these two results. In our experiments, we installed speakers at intervals of 10 m along a corridor and estimated the smartphone position at several places. From these results, we obtained 90th percentile errors of less than 0.224 m for 2-D positioning. We found that multipaths from the side wall were causing the positioning error in passive acoustic sensing, and the variance of the positioning error using the top microphone which was omnidirectional was smaller than the bottom one. When we introduced a weighting based on the result of the active acoustic sensing and the difference in the performance between microphones, the 90th percentile errors were reduced to less than 0.134 m.**

***Keywords*—active and passive acoustic sensing, TDoA, smartphone, microphone**

I. Introduction

In the development of the "internet of things," positional information is often required. The Global Positioning System (GPS) exists as a general position-estimation system, but GPS signals do not reach indoors. There remain many scenarios where we must rely on brochures and signs in shopping malls, basements, or museums. Therefore, there is high demand for an effective indoor position-recognition technology.

One indoor position-recognition approach involves acoustic signals. These methods usually adopt Time Difference of Arrival (TDoA) multilateration instead of using Time of Arrival (ToA) because it is difficult accurately to calibrate the time between the speakers transmitting the signals and the receiving devices. Position estimation using acoustic signals has the advantage that position-estimation accuracy is higher than for Received Signal Strength Indication (RSSI) methods, which involve radio field intensity. Moreover, if we can use existing speaker installations, we can estimate the position without additional equipment. However, at least three speakers are required for 2-D positioning using the TDoA approach. This can make position estimation difficult in some situations, such as corridors, because of the insufficient number of speakers.

978-1-5386-5635-8/18/$31.00

Therefore, we propose an indoor location-estimation method based on acoustic sensing that involves only two installed speakers and a smartphone. Speakers are installed on a straight line at a regular interval in corridors of many buildings, giving a big advantage to a system able to measure locations using only two installed speakers.

We can estimate the smartphone's location by using the acoustic signals from two adjacent speakers (passive acoustic sensing) and reflected acoustic signals transmitted from the smartphone (active acoustic sensing). In this paper, we describe the construction of position-estimation models using the two microphones built into a smartphone. We have demonstrated the effectiveness of our proposed method through experimental evaluation, achieving 90th percentile errors of less than 0.134 m.

II. Related Work

The Active Bat system[1] is an ultrasonic position-recognition system that uses the ToA method. Several ultrasonic receivers are installed indoors, which estimate the position of an ultrasonic transmitter called the Active Bat tag. Hazas et al.[1] use broadband ultrasonic signals and reported a 95th percentile position error of less than 0.05 m for 3-D positioning with a time-synchronized transmitter and receiver. In Cricket[2], beacons are installed that transmit RF signals and ultrasonic signals indoors, estimating the position of the receiver by ToA of reception for the two signals. This method achieved a positional accuracy of 0.01–0.02 m in a room with a maximum volume of $1,000$ m^3. This system has been extended to Cricket Compass[3], which measures the angle of the target by Angle of Arrival (AoA). The ambiguity with respect to 2π periods in the angle measurement is avoided by using five ultrasonic receiving elements in an L shape at the receiver.

Because achieving fast and accurate time synchronization between transmitter and receiver using a smartphone remains difficult, any positioning system based on the ToA method will also face problems. For this reason, position recognition by the TDoA method is being applied widely. Furthermore, for high audible frequencies, the frequency characteristics of the microphone and the speaker are poor, and the sound pressure level can be low or unstable. Therefore, it is difficult to match the positional performance of ultrasonic position-recognition systems such as [1] or [2]

In position-recognition systems based on acoustic signals, frequency chirp signals are often used. If the frequency scan is accelerated by using a chirp signal, frequency-modulated sidebands are generated, making signal transmission difficult. Achieving simultaneously both high accuracy and a short measurement time using a chirp signal is difficult. ASSIST[4] transmits a chirp signal from a smartphone and calculates a 2-D position using the TDoA method. If an indoor installed microphone is unavailable, this system can estimate the position using the acceleration sensor built into a smartphone and a Kalman filter. With position recognition using acoustic signals, we may hear noise caused by the speaker transient response, for example, even when using high audible frequencies that people find hard to hear. Lazik et al.[5] propose a method for solving this problem by applying fade-in and fade-out to the chirp signal. Liu et al.[6] improve the position-recognition accuracy by using the positional information from Wi-Fi together with the distance information between smartphones from BeepBeep[7] using a chirp signal. In WalkieLokie[8], a tag is installed that can transmit/receive audio signals to/from users. The system calculates the relative position of users from the phase shift caused by the Doppler shift.

Recently, many systems that use the echoes of acoustic signals have been proposed. BatMapper[9] uses echoes from smartphones and can generate a map of a floor. This system can detect doors during the map creation by using different optimum acoustic signals for each of the top and bottom microphones. Similarly, BatTracker[10] can track the relative position of the smartphone. Oliver et al.[11] propose a method to measure the occupancy rate in the room by variations in the frequency response of the high-frequency chirp signal transmitted from a tweeter in the center of the room.

For these position-recognition methods based on acoustic signals, we need dedicated devices and three or more speakers. However, in an actual living environment, additional equipment should be avoided wherever possible. One advantage of our proposed method is that position estimation is possible without additional equipment. It uses the reverberation sound of the acoustic signal transmitted from the smartphone in addition to the acoustic signals from two existing installed speakers.

III. Proposed System

A. Problem Settings

With TDoA multilateration, at least three speakers are required for 2-D positioning. Although more than two speakers are installed in a corridor of a building, signals emitted from far-away speakers are not stably detected at a receiver device and the number of available speakers for positioning becomes less than three. Our approach to solving this problem is to also use the echoes from a smartphone's speakers. We, therefore, propose a position-estimation model that uses two of the existing installed speakers, the smartphone's speakers, and its microphones.

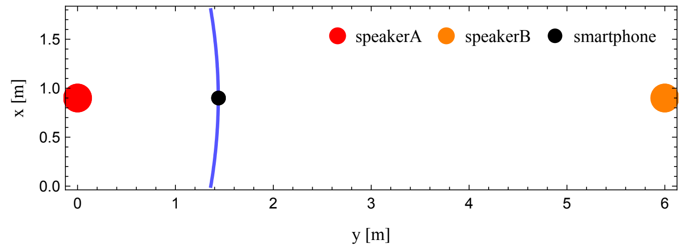

Fig. 1. TDoA using two speakers

B. Passive Acoustic Sensing

1) Position-Calculation Method: For the position-calculation method, we use TDoA multilateration, which does not require time synchronization of the transmitter and receiver. The position of the receiver is calculated from the distance differences between multiple transmitters and receivers. If the 2-D position of a transmitter i is $(x_i, y_i, z_i), (i = 1, .., 3)$, the 2-D position of a receiver (x, y) is given as:

$$\begin{cases} \sqrt{(x-x_2)^2+(y-y_2)^2+(z_c-z_2)^2} \\ \quad -\sqrt{(x-x_1)^2+(y-y_1)^2+(z_c-z_1)^2} = c(t_2-t_1), \\ \sqrt{(x-x_3)^2+(y-y_3)^2+(z_c-z_3)^2} \\ \quad -\sqrt{(x-x_1)^2+(y-y_1)^2+(z_c-z_1)^2} = c(t_3-t_1), \end{cases} \tag{1}$$

where c is the sound velocity and t_i is the arrival time at the receiver of the acoustic signal from transmitter i. z_c is z coordinate of the smartphone.

In this paper, we assume the use of two speakers, leading to a curved locus of position in 2-D space, as shown in Fig.1.

2) Signal Detection Method: In TDoA mulilateration, the detection accuracy of the reception time affects the positional accuracy. Therefore, we adopt the Phase Accordance Method (PAM)[12], devised by our research group, which enables highly accurate reception-time detection with narrowband signals. In PAM, we use a signal called "sync pattern" (see Fig.2), which is the sum of two sine waves of different frequencies, and which can be expressed as:

$$\begin{aligned} s(t) &= a_1\sin(\omega_1 t+\phi_1)+a_2\sin(\omega_2 t+\phi_2) \\ &= a_1\sin(2\pi f_1 t+\phi_1)+a_2\sin(2\pi f_2 t+\phi_2), \end{aligned} \tag{2}$$

where a_i is the amplitude, ω_i is the angular frequency, f_i is the frequency, and ϕ_i is the initial phase $(i = 1, 2)$. We can find the outline of sync pattern by correlating the received data with the transmitted signal. We can calculate the phases of two sinusoidal waves by applying a rectangular window to the received sync pattern and thereby calculate the phase difference. As a result, we obtain the phase difference of the waves, as shown in Fig.2. There is only one point where the phase difference becomes zero, which coincides with the center of the sync pattern. Therefore, we can determine the arrival time accurately by specifying this point as the signal transmission time.

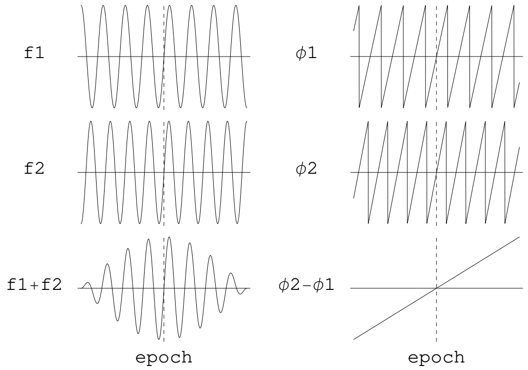

Fig. 2. Sync pattern and epoch

There are two methods of signal communication using PAM[12], Frequency-Division Multiplexing (FDM) and Time-Division Multiplexing (TDM). FDM-PAM can receive all signals at once because it uses four orthogonal frequency pairs, including a human-inaudible band: $(14.75, 15.25)$kHz, $(15.75, 16.25)$kHz, $(16.75, 17.25)$kHz, and $(17.75, 18.25)$kHz. FDM-PAM can, therefore, reduce the measurement time. In addition, highly accurate positioning is possible because errors caused by movement are avoided. However, the measurable area for positioning is very small, as described in [13]. For this reason, we prefer to use TDM-PAM, to enable a larger measurable area. Here, we use two frequency pairs, $(14.75, 15.25)$kHz and $(16.75, 17.25)$kHz, to distinguish the speakers. Therefore, the length of sync pattern we use in the experiment is 2 ms.

3) Reduction of Multipaths: If there is a Line of Sight (LOS), we think that the time with the highest correlation value is the direct path of the sync pattern. However, the correlation value for a multipath signal can sometimes exceed that for the direct path's signal because of the directivity of the microphones (see Fig.3). Therefore, we consider the first path with the first peak as a direct path and propose a threshold-based algorithm to find first path. If the arrival time at which the correlation is maximum is t_p, the time window for obtaining the threshold is T_t. We set the threshold value based on the energy E including the received multipath signals. We set the threshold value η as the product of the time-averaged value of E and an experimentally obtained coefficient m. This gives the threshold expression:

$$E = \sum_{t=t_p}^{T_t} \frac{1}{2}(a[t] + a[t+1])\Delta t \tag{3}$$

$$\eta = \frac{Em}{T_t}, \tag{4}$$

where $a[t]$ is the time-series data of the envelope obtained from the correlation between the transmission signal and the received signal and Δt is the sampling time interval.

Assuming that the direct signal reception occurs before t_p, we can then recognize the first peak value exceeding η in the range T_d as that of the first-path signal.

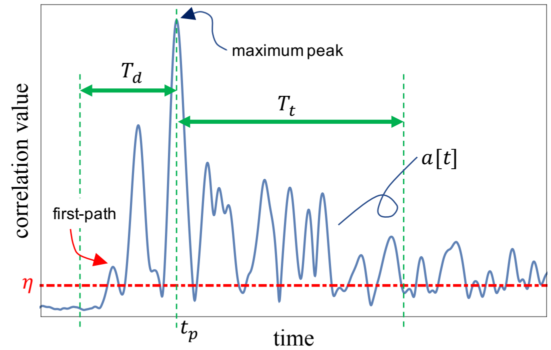

Fig. 3. First-path and multipath signals in passive acoustic sensing

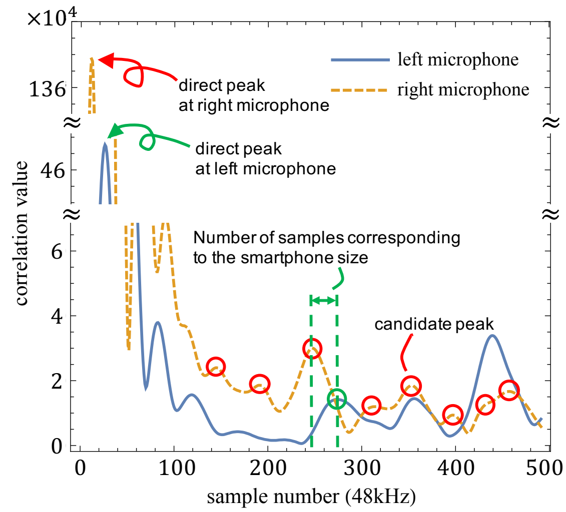

Fig. 4. Echoes obtained by two microphones in active acoustic sensing

C. Active Acoustic Sensing

We now propose a method to calculate the distance between a smartphone and the side walls of a room or corridor by using the echoes of active acoustic signals transmitted from the smartphone. We use two microphones built into the smartphone. The arrival time of the echoes obtained by the two microphones (see Fig. 4) depends on the distances to the side walls. We can estimate the distance between the smartphone and the sidewall surfaces by considering their geometric relationship. The relationship assumes that the smartphone axis is orthogonal to the side walls (see Fig5). In our experiments, we pointed the smartphone's top microphone to the righthand wall and the bottom one to the left.

1) Design of Transmission Signal: We use a chirp as a transmission signal, which is also used in sonar and radar. A linear chirp can be expressed as:

$$c(t) = \sin(\phi_0 + 2\pi(f_0 t + \frac{k}{2}t^2)) \tag{5}$$

$$k = \frac{f_1 - f_0}{T},$$

where ϕ_0 is the initial phase, f_0 is the start frequency, f_1 is the end frequency, and T is the time to sweep from f_0 to f_1. The signal-to-noise ratio is improved by applying a Hanning window to the transmitted chirp signal. We set $\phi_0 = 0$, $f_0 = 8$kHz, $f_1 = 9$kHz, and $T = 1$ms to avoid the frequency bands containing environmental noise, human voices, and passive acoustic-sensing signals. We use the smartphone's top and bottom speakers for signal transmission. In our experiments, we held the smartphone sideways, using the top speaker as the righthand speaker and the bottom speaker as the lefthand speaker. Instead of simultaneous transmission from the two speakers, the signals are transmitted alternately every 500 ms to minimize interference between the echoes.

2) Preprocessing: We preprocess the acoustic data obtained by each microphone to improve the ranging accuracy. To reduce the noise caused by environmental sounds and passive acoustic-sensing signals, we apply a bandpass filter corresponding to the frequency band of the received data. We obtain the envelope for the peak values of the echoes by cross-correlation between the transmitted signal and the received-signal and apply smoothing to exclude outliers. We need to know the signal transmission time, which can be found from the time stamp. However, it is possible for large errors to occur in the time stamp. We, therefore, use the initial direct sound, which travels to the microphone through the smartphone itself whenever we transmit a signal. This is much more reliable than the time stamp because the signal travels rapidly through the smartphone. Therefore, we can use the time of the first peak in the echo signal as the signal transmission time. In our experiments, we set the time window as 10 ms.

3) Distance Estimation: We focus on the arrival time difference of echoes between microphones to estimate the distance d between the smartphone and the side walls. To calculate the distance d between the smartphone and a side wall, we need to obtain the reciprocating distance r from the smartphone's speaker to smartphone's microphones. If t_s is the time when the signal is transmitted from the speaker, t_f is the time when echoes are received at the microphone, with a sound velocity of c, then:

$$r = c(t_f - t_s) \tag{6}$$

We can obtain the signal transmission time t_s by the arrival time of the first peak value. Therefore, we need to specify the reception time of the peak value corresponding to the signal reception time t_f to obtain the reciprocating distance r to the side wall.

We now describe how to find t_f. Let $t_{R,i} (i = 1, ..., 8)$ be the reception times for the first eight peak values obtained by the right microphone (see Fig.4). Let $t_{L.j} (j = 1, ..., 12)$ be the reception times for the first 12 peak values obtained by the left microphone. The right microphone has better sensitivity with respect to noise canceling than the left microphone, and receives many multipath signals and noise, so we change the order of selection. Let $r_{R,i}$ and $r_{L,j}$ be the estimated distances from the smartphone to the side walls with respect to the reception times $t_{R,i}$ and $t_{L,j}$. By using the geometric relationship between the smartphone and the side walls (see Fig5), if the reception time $t_{R,i}$ is the echo from the righthand wall, then:

$$\begin{aligned} -l_p + \varepsilon_R &= r_{R,i} - r_{L,j} \\ &= c(t_{R,i} - t_s) - c(t_{L,j} - t_s) \\ &= c(t_{R,i} - t_{L,j}), \end{aligned} \tag{7}$$

where l_p is the width of the smartphone used and ε_R is a noise-related value. This situation is illustrated in Fig.4. We can calculate $d_{R,i}$ (see Fig.5(a)) by $r_{R,i}$ that satisfies the above conditions and can express as:

$$d_{R,i} = \begin{cases} \frac{1}{2} r_{R,i} & \text{if the signal is transmitted} \\ & \quad \text{from the right speaker,} \\ \frac{1}{2}(r_{R,i} - l_p) & \text{if the signal is transmitted} \\ & \quad \text{from the left speaker.} \end{cases} \tag{8}$$

If the reception time $t_{L,j}$ refers to the echo from the left side wall:

$$\begin{aligned} l_p + \varepsilon_L &= r_{R,i} - r_{L,j} \\ &= c(t_{R,i} - t_s) - c(t_{L,j} - t_s) \\ &= c(t_{R,i} - t_{L,j}), \end{aligned} \tag{9}$$

where ε_L is a noise-related value. Therefore, $d_{L,j}$ can be expressed as:

$$d_{L,j} = \begin{cases} \frac{1}{2}(r_{L,j} - l_p) & \text{if the signal is transmitted} \\ & \quad \text{from the right speaker,} \\ \frac{1}{2} r_{L,j} & \text{if the signal is transmitted} \\ & \quad \text{from the left speaker.} \end{cases} \tag{10}$$

From $d_{R,i}$ and $d_{L,j}$ that satisfy the above conditions, we select the final candidate by using the width w_c of the corridor. To improve the reliability, we select the two most recent candidates of $d_{R,i}^{(R)}$ and $d_{L,j}^{(R)}$ for the signal transmitted from the right speaker and $d_{R,i}^{(L)}$ and $d_{L,j}^{(L)}$ for the signal transmitted from the left speaker, in chronological order. Then, the $d_{R,i}$ and $d_{L,j}$ satisfying one of the following expressions is the combination of the estimated distance $d_{R,i}$, $d_{L,j}$:

$$w_c - v \leq d_{R,i}^{(R)} + d_{L,j}^{(R)} + l_p \leq w_c + v, \tag{11}$$

$$w_c - v \leq d_{R,i}^{(R)} + d_{L,j}^{(L)} + l_p \leq w_c + v, \tag{12}$$

$$w_c - v \leq d_{R,i}^{(L)} + d_{L,j}^{(R)} + l_p \leq w_c + v, \tag{13}$$

$$w_c - v \leq d_{R,i}^{(L)} + d_{L,j}^{(L)} + l_p \leq w_c + v, \tag{14}$$

where v is a noise-related value.

If we still have several candidates, we choose the one for which $|\varepsilon_R| + |\varepsilon_L|$ is the smallest.

4) Removing Noise Caused by Reverberation in the Smartphone: When the acoustic signal is transmitted from the internal speaker of the smartphone, there are regular echoes generated inside the smartphone (see Fig.6). Reverberation generated internally has a very large amplitude, and this may cause errors in the distance measurement. Therefore, we

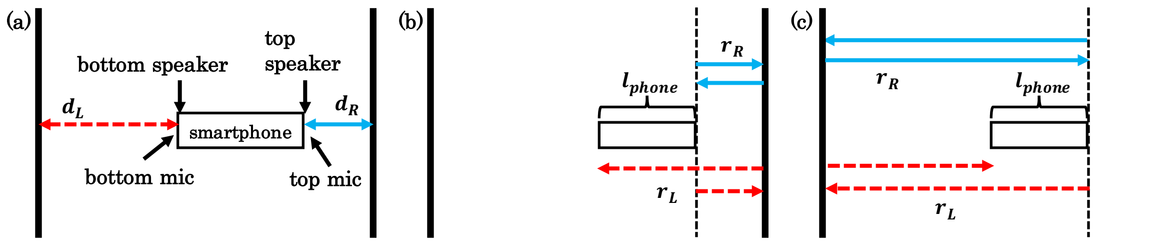

Fig. 5. Geometric relationship between a smartphone and side walls: (a) positioning of the smartphone, (b) echoes from the right wall, (c) echoes from the left wall

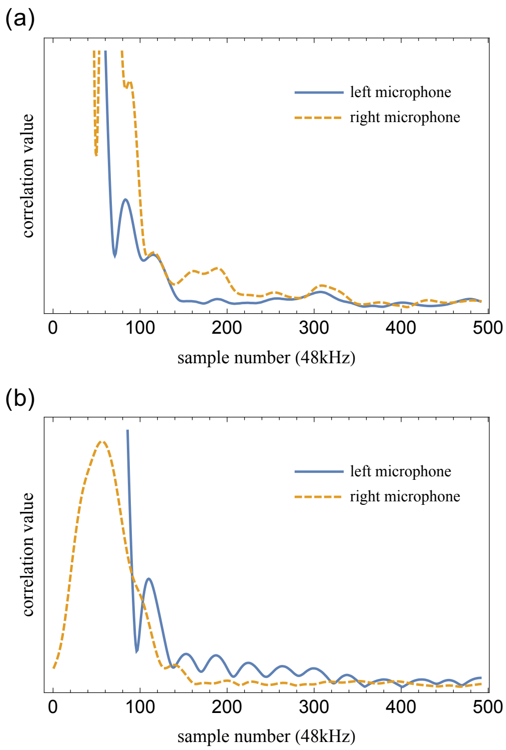

Fig. 6. Reverberation sound inside the smartphone: (a) signal transmitted by right speaker, (b) signal transmitted by left speaker

exclude as noise 130 samples at 48 kHz that are considered to be internal echoes. This can be a problem if the distance to the side wall is very short because the wall echo is then included in this "noise." However, when we use two speakers, we can measure distances to the side wall of more than 0.3 m, for a smartphone width of 0.145 m and a sound velocity of 340 m/s. This distance is considered to be a reasonable limit when walking along the edge of the corridor.

D. Position-Estimation Model Using Acoustic Sensing

We propose a position-estimation method that uses both active and passive acoustic sensing. The position-estimation model involves two microphones in a smartphone. Consider two ceiling-speaker coordinate sets $S_A(x_A, y_A, z_A)$ and $S_B(x_B, y_B, z_B)$, the smartphone coordinate set $P(x, y, z)$, and coordinate sets for the right and left microphones in the

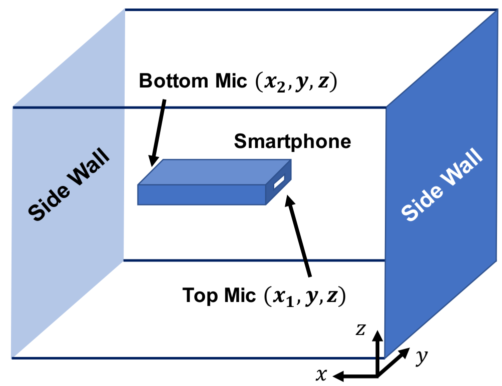

Fig. 7. Configuration of model

smartphone $M_R(x_1, y, z)$ and $M_L(x_2, y, z)$, respectively (see Fig.7). Using both passive and active acoustic sensing, we have:

$$\begin{cases} \sqrt{(x_1-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x_1-x_B)^2+(y-y_B)^2+(z-z_B)^2} = c * t_1 \\ \sqrt{(x_2-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x_2-x_B)^2+(y-y_B)^2+(z-z_B)^2} = c * t_2 \\ x_1 = x - \frac{1}{2}l_p \\ x_2 = x + \frac{1}{2}l_p \\ x = x_e, \end{cases} \tag{15}$$

where t_1 and t_2 are the TDoA values for passive acoustic sensing obtained by the right and left microphones, respectively, and x_e is the x coordinate of the smartphone obtained by active acoustic sensing. Equation (15) can be transformed to:

$$\begin{cases} \sqrt{(x-\frac{1}{2}l_p-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x-\frac{1}{2}l_p-x_B)^2+(y-y_B)^2+(z-z_B)^2} = c * t_1 \\ \sqrt{(x+\frac{1}{2}l_p-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x+\frac{1}{2}l_p-x_B)^2+(y-y_B)^2+(z-z_B)^2} = c * t_2 \\ x = x_e \end{cases} \tag{16}$$

We solve equation (17) using the least-squares method:

$$J = \sum_{i=1}^{3} f_i(x, y, z)^2, \tag{17}$$

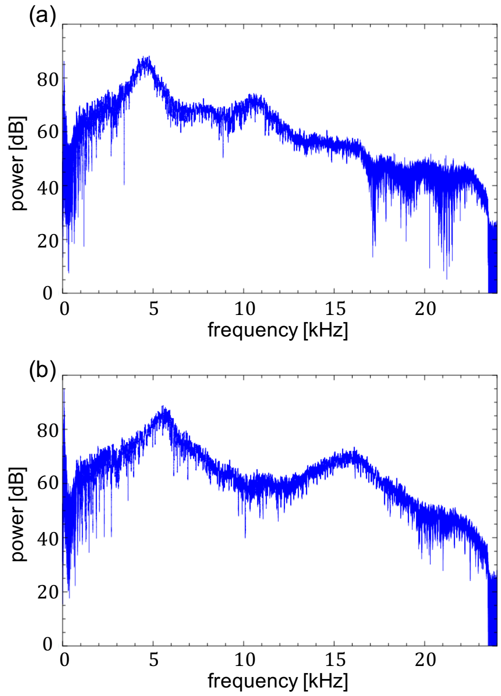

Fig. 8. Frequency characteristics: (a) right microphone (b) left microphone

where:

$$\begin{cases} f_1(x,y,z) = \sqrt{(x-\frac{1}{2}l_p-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x-\frac{1}{2}l_p-x_B)^2+(y-y_B)^2+(z-z_B)^2} - c*t_1 \\ f_2(x,y,z) = \sqrt{(x+\frac{1}{2}l_p-x_A)^2+(y-y_A)^2+(z-z_A)^2} \\ \quad -\sqrt{(x+\frac{1}{2}l_p-x_B)^2+(y-y_B)^2+(z-z_B)^2} - c*t_2 \\ f_3(x,y,z) = x - x_e \end{cases}$$

IV. Experimental Evaluation

A. Preliminary Experiment: Survey of the Frequency Characteristics of Microphones

We investigated the frequency characteristics of the right and left microphones of a smartphone (SONY SO-03F). We used the linear chirp signal that transitions from 0 to 23.5 kHz in 10 s using a Fostex FT200D as the speaker. Figs.8(a) and 8(b) show the frequency characteristics for each microphone. The results demonstrate that the microphones used in our experiments have a sufficiently flat frequency response in the frequency band used.

B. Experiment: Positioning Experiments in a Real Environment

1) Experiment Setting: To confirm the effectiveness of the proposed method, positioning experiments were conducted with two speakers at nine locations in a corridor (see Fig.9). The distance between the speakers was 10 m. The z coordinate of the speakers was 2.2 m, and the z coordinate of the smartphone at positions P1–P9 was 1.1 m. We used a Fostex FT200D for the transmission speaker, a SONY SO-03F for

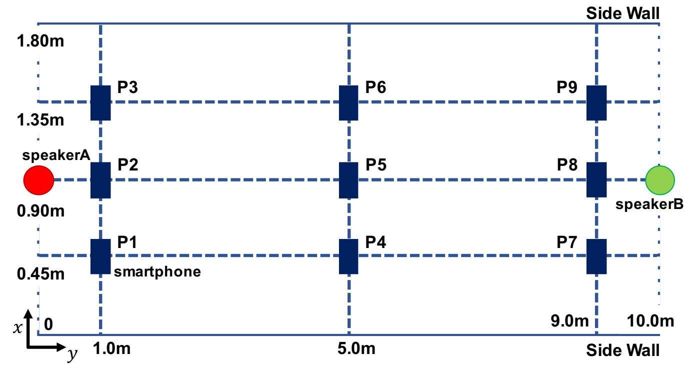

Fig. 9. Experimental environment

the smartphone, an NF WF1948 as the signal generator, and our own amplifier as the speaker driver. We set 48 kHz as the sampling rate for audio recording. To investigate the effectiveness of the proposed method, the following three experiments were conducted:

1. Position measurement based on passive acoustic sensing
2. Distance measurement based on active acoustic sensing
3. Position estimation by integrating 1. and 2.

In Experiment 1, we investigated the minimum distance between the true position and the TDoA curves obtained by passive acoustic sensing. In Experiment 2, we investigated the distance measurement error between the true and estimated positions by active acoustic sensing. We then evaluated the overall effectiveness of our proposed method in Experiment 3.

2) Experiment 1: Position Measurement Based on Passive Acoustic Sensing: We performed a 2-D positioning experiment using only passive acoustic sensing to check the usefulness of the position-estimation model. We set $z = 1.1$ m and estimated the locus of 2-D positions for the smartphone 100 times by using passive acoustic sensing. Fig.10 shows the minimum distance between the true position and the TDoA curves obtained by 2-D passive acoustic sensing. We found that the 90th percentile maximum error was 0.280 m for the right microphone and 0.386 m for the left microphone.

3) Experiment 2: Distance Measurement Based on Active Acoustic Sensing: We investigated the distance error between the true x coordinate and the estimated x coordinate by our proposed active-acoustic-sensing method. We estimated the distance 100 times at three locations: $x = 0.45$ m, $x = 0.90$ m, $x = 1.35$ m. Fig.11 shows the cumulative error function for these locations. We found that the maximum error was 0.117 m and the 90th percentile maximum error was 0.0803 m.

4) Experiment 3: Integrated Position Estimation: To evaluate the overall performance of our proposed method, we conducted a 2-D position-estimation experiment 100 times. We set $z = 1.1$ m, as in Experiment 1. Fig.12 shows the cumulative error function of the estimated 2-D position. We found that the 90th percentile error was less than 0.224 m.

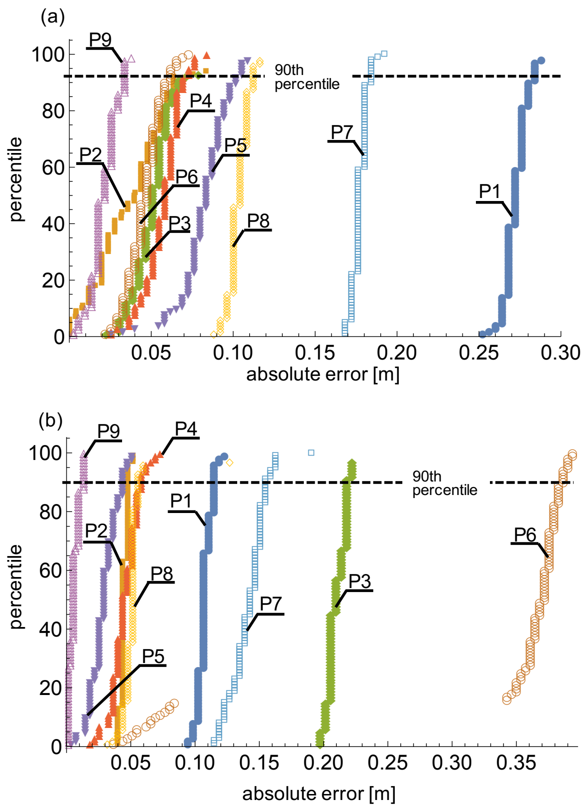

Fig. 10. Cumulative error function for passive acoustic sensing: (a) right microphone, (b) left microphone

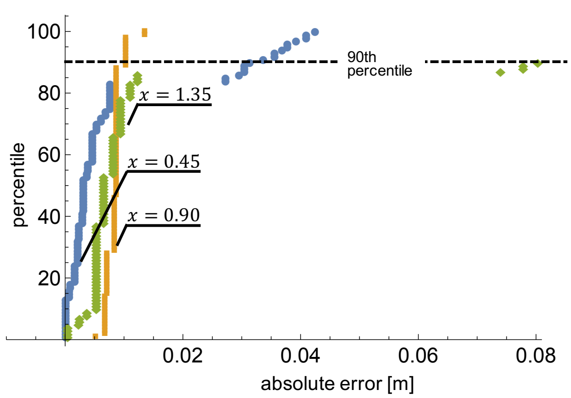

Fig. 11. Cumulative error function for active acoustic sensing

V. Discussion

1) Experiment 1: There are circumstances for which large errors in 2-D positioning can occur. We know that the performance of the microphone close to the side wall can be poor (see Fig.10). For example, for P1 or P7 close to the righthand wall, the position error for the right microphone is large. For P6 near the lefthand wall, the positioning error of the left microphone is large. We consider that the likely cause is a multipath issue. When the second-path signal arrives within 2 ms $\times$ 340 m/s $=$ 0.68 m of the first-path signal, this signal

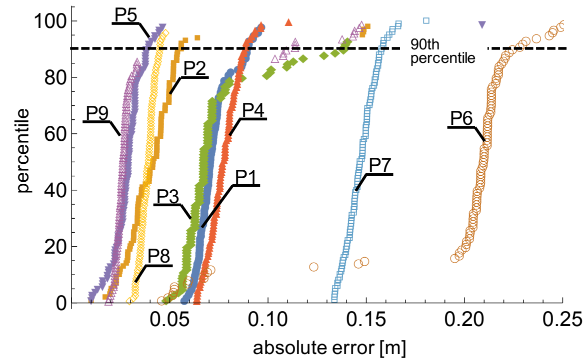

Fig. 12. Cumulative error function for the proposed method

appears to merge with it. As a result, the first-path peak cannot be identified clearly, leading to a significant positioning error.

We also found that we could not calculate a 3D position of smartphone using equations (15), although they can theoretically give the solution. Here, the cause is considered to involve both multipath and LOS issues. We hold the smartphone sideways to enable active acoustic sensing. The speaker is then positioned perpendicular to the microphone direction for the smartphone. In our proposed method, it is impossible to always achieve LOS with both microphones. Therefore, we consider that the microphone might be receiving a diffracted direct-path signal, causing a reception-time error. Note that the proposed model equation is very sensitive to reception-time errors because the baseline length for the two microphones is very short. Therefore, we consider that we cannot estimate the 3-D position at present.

2) Experiment 3: The large positioning errors in Experiment 3 were mainly caused by multipath signals, as described above. We found that the reliability of the right microphone is lower if the smartphone is on the righthand side of the corridor and the reliability of the left microphone is lower if it is on the lefthand side. Therefore, to minimize this effect, we propose to use the weighted least-squares method with the results of the active acoustic sensing. Also, we confirmed that the right microphone had smaller variance of the positioning error than the left microphone. We consider the cause is the fact that the right microphone for noise cancellation is omnidirectional and the left microphone for calling is unidirectional. Therefore, we give large weight to the right microphone than to the left one. Using these method, we obtain:

$$J = \sum_{i=1}^{3} \tilde{w_i} f_i(x,y,z)^2 \tag{18}$$

where

$$\boldsymbol{w} = \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} = \begin{pmatrix} \alpha(\frac{x_e}{w_c})^2 \\ (1-\alpha)(1-\frac{x_e}{w_c})^2 \\ \frac{1}{2}(w_1+w_2) \end{pmatrix}, \tag{19}$$

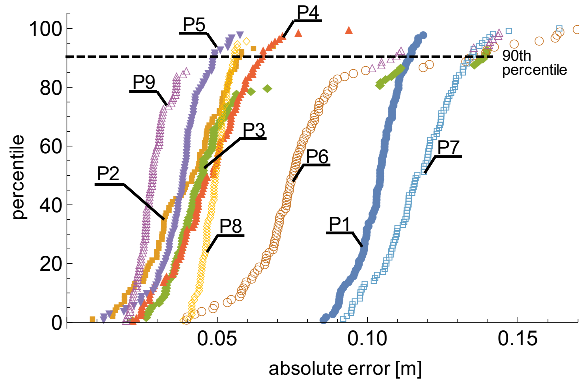

Fig. 13. Cumulative error function for the proposed method with weighted active acoustic sensing

$$\tilde{\boldsymbol{w}} = \begin{pmatrix} \tilde{w}_1 \\ \tilde{w}_2 \\ \tilde{w}_3 \end{pmatrix} = \frac{1}{w_1 + w_2 + w_3} \begin{pmatrix} w_1 \\ w_2 \\ w_3 \end{pmatrix} \tag{20}$$

In this method, we assume that $\frac{x_e}{w_c} = 0$ on the right and $\frac{x_e}{w_c} = 1$ on the left in a corridor. We, also, set $\alpha = 0.6$ based on the result of the several experiments. Fig.13 shows the cumulative error function for the proposed method when weighted active acoustic sensing is used. We found that the 90th percentile error was now less than 0.134 m and the performance had greatly improved.

VI. Conclusions and Future Work

In this paper, we have described our proposed position-estimation method, which combines passive acoustic sensing using two existing installed speakers and active acoustic sensing using smartphone loudspeakers. We installed two speakers at intervals of 10 m in a corridor and conducted position-estimation experiments at nine different places. The results demonstrated that 90th percentile errors of less than 0.134m for 2-D position estimation could be achieved. In future work, we will conduct experiments using different models of smartphones and aim to improve the performance of the positioning method. We will also address the challenge of 3-D position estimation and plan to develop applications for indoor navigation using smartphones.

Acknowledgment

This work was supported by JSPS KAKENHI Grant Number 16H03070 and 17K19952.

References

[1] Mike Hazas and Andy Hopper. Broadband ultrasonic location systems for improved indoor positioning. *IEEE Transactions on mobile Computing*, 5(5):536–547, 2006.

[2] Nissanka B Priyantha, Anit Chakraborty, and Hari Balakrishnan. The cricket location-support system. In *Proceedings of the 6th annual international conference on Mobile computing and networking*, pages 32–43. ACM, 2000.

[3] Nissanka B Priyantha, Allen KL Miu, Hari Balakrishnan, and Seth Teller. The cricket compass for context-aware mobile applications. In *Proceedings of the 7th annual international conference on Mobile computing and networking*, pages 1–14. ACM, 2001.

[4] Fabian Höflinger, Rui Zhang, Joachim Hoppe, Amir Bannoura, Leonhard M Reindl, Johannes Wendeberg, Manuel Bührer, and Christian Schindelhauer. Acoustic self-calibrating system for indoor smartphone tracking (assist). In *Indoor Positioning and Indoor Navigation (IPIN), 2012 International Conference on*, pages 1–9. IEEE, 2012.

[5] Patrick Lazik and Anthony Rowe. Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In *Proceedings of the 10th ACM Conference on Embedded Network Sensor Systems*, pages 99–112. ACM, 2012.

[6] Hongbo Liu, Yu Gan, Jie Yang, Simon Sidhom, Yan Wang, Yingying Chen, and Fan Ye. Push the limit of wifi based localization for smartphones. In *Proceedings of the 18th annual international conference on Mobile computing and networking*, pages 305–316. ACM, 2012.

[7] Chunyi Peng, Guobin Shen, Yongguang Zhang, Yanlin Li, and Kun Tan. Beepbeep: a high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of the 5th international conference on Embedded networked sensor systems*, pages 1–14. ACM, 2007.

[8] Wenchao Huang, Xiang-Yang Li, Yan Xiong, Panlong Yang, Yiqing Hu, Xufei Mao, Fuyou Miao, Baohua Zhao, and Jumin Zhao. Walkielokie: Sensing relative positions of surrounding presenters by acoustic signals. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 439–450. ACM, 2016.

[9] Bing Zhou, Mohammed Elbadry, Ruipeng Gao, and Fan Ye. Batmapper: Acoustic sensing based indoor floor plan construction using smartphones. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services*, pages 42–55. ACM, 2017.

[10] Bing Zhou, Mohammed Elbadry, Ruipeng Gao, and Fan Ye. Battracker: High precision infrastructure-free mobile device tracking in indoor environments. 2017.

[11] Oliver Shih and Anthony Rowe. Occupancy estimation using ultrasonic chirps. In *Proceedings of the ACM/IEEE Sixth International Conference on Cyber-Physical Systems*, pages 149–158. ACM, 2015.

[12] Hiromichi Hashizume, Ayumu Kaneko, Yusuke Sugano, Koji Yatani, and Masanori Sugimoto. Fast and accurate positioning technique using ultrasonic phase accordance method. In *TENCON 2005 2005 IEEE Region 10*, pages 1–6. IEEE, 2005.

[13] Masanari Nakamura, Takayuki Akiyama, Masanori Sugimoto, and Hiromichi Hashizume. 3d fdm-pam: rapid and precise indoor 3d localization using acoustic signal for smartphone. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing: Adjunct Publication*, pages 123–126. ACM, 2014.