Expanding the Positioning Area for Acoustic Localization Using COTS Mobile Devices

Takumi Suzaki[1(✉)], Masanari Nakamura[1], Hiroaki Murakami[2], Hiroki Watanabe[1], Hiromichi Hashizume[3], and Masanori Sugimoto[1]

[1] Hokkaido University, Sapporo, Japan
{tsuzaki,masanari,hiroki.watanabe,sugi}@ist.hokudai.ac.jp
[2] The University of Tokyo, Tokyo, Japan
murakami@akg.t.u-tokyo.ac.jp
[3] National Institute of Informatics, Tokyo, Japan
has@nii.ac.jp

Abstract. In this paper, we propose a novel acoustic localization system using commercial off-the-shelf (COTS) mobile devices. Acoustic-based systems have advantages in terms of accuracy and cost. However, the measurable positioning area is limited because of the signal attenuation and the poor performance of microphones embedded in COTS mobile devices. Our system leverages a transmission scheme that combines time-division multiple-access (TDMA) and frequency-division multiple-access (FDMA) techniques to address the limitation. In the proposed approach, each speaker transmits different band chirps in a predefined sequence to mitigate multiple-access interference. A COTS device receives modulated signals via a built-in microphone. We exploit the received signals and estimate the position by calculating time difference of arrival (TDoA). We were able to reduce the error to a 90th-percentile error of 46.26 cm at a measurement point that could not be estimated by FDMA-based positioning. The experiment results show that our system is more accurate and has a larger area of positioning capability compared with FDMA-based positioning.

Keywords: Chirp · Acoustic indoor localization · TDoA · Smartphone · FDMA and TDMA · Sensing

1 Introduction

It has been reported [4] that people spend 88.9% of their days indoors on average and have many experiences that involve information about their indoor location. In addition, because of the COVID-19 pandemic, it is necessary to have a system that monitors whether sufficient social distancing is maintained. By 2025, the market value of indoor location information is expected to exceed 43 billion dollars [15], proving that there is a large demand for effective indoor localization

T. Hara and H. Yamaguchi (Eds.): MobiQuitous 2021, LNICST 419, pp. 422–437, 2022.
https://doi.org/10.1007/978-3-030-94822-1_23

technology. Unfortunately, it is difficult to use a global positioning system (GPS) indoors because the weak signals do not easily penetrate building walls [25]. Therefore, various indoor localization systems have been proposed, such as Wi-Fi [2], Bluetooth [7], and UWB [14]. However, a standard system for indoor localization systems has not yet been proposed because current systems need improvements regarding accuracy and implementation cost.

This paper describes an indoor localization system using acoustic signals. Acoustic signal-based localization technology leverages microphone sensors in mobile devices to capture acoustic signals transmitted by sound sources and to estimate user locations. Acoustic-based systems have been shown to achieve high localization accuracy. Furthermore, because microphones are embedded in COTS mobile devices and speakers are installed in an indoor environment, they can be used to estimate the positions without additional hardware.

Regarding positioning performance, the precision of positioning mainly depends on the transmitted signal. In general, the wider the bandwidth of the signal, the more precise TDoA estimation becomes. However, from the perspective of scalability, the bandwidth should be as narrow as possible. Similarly, the longer the length of the signal, the more precise TDoA estimation becomes due to signal-to-noise ratio (SNR). However, from the perspective of the positioning update rate, the length of the signal should be as short as possible. Given these facts, a signal with a narrow bandwidth and as short a length as possible is suitable for many environments.

However, a system with a narrow bandwidth and short signal length, as described, has a limited positioning area. This is because it is difficult to detect signals over a large area due to the low performance of the microphones embedded in COTS devices (described later).

In this paper, we propose a transmission scheme that combines TDMA and FDMA techniques to reduce the limitation of positioning area. Robust positioning is achieved by applying coherent averaging to the received signals.

To evaluate our proposed method, we experimented with three speakers performing two-dimensional positioning estimation using TDoA. To the best of our knowledge, the proposed method in this paper has not been reported in previous literature. Thus, this is the first attempt to show how it can effectively alleviate problems regarding acoustic positioning in practical situations.

The main contributions of this paper are as follows:

- We propose a transmission scheme that combines TDMA and FDMA techniques to expand the positioning area for COTS mobile devices.
- We utilize a coherent averaging for received signals to improve the range resolution.
- We investigate the delay that occurs when high-frequency signals are received by COTS devices.

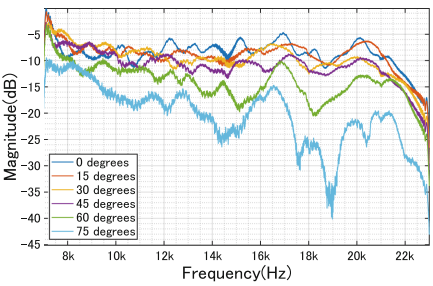


Fig. 1. Frequency response by arrival angle (Samsung Galaxy S10 plus)

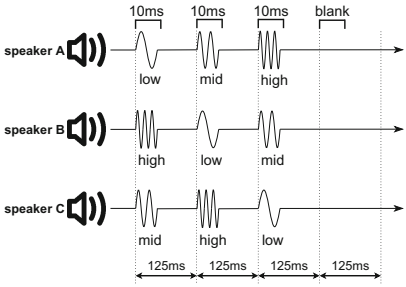


Fig. 2. Outline of the transmission scheme

2 Proposed Method

2.1 Chirp Signals

We use a chirp signal, which is widely used in sonar and radar systems, as the transmission signal. It has a characteristic of the frequency increasing linearly with time (called up-chirp). A linear chirp can be expressed as:

$$s(t) = \sin\{2\pi(f_0 t + \frac{k}{2}t^2)\}, \tag{1}$$

$$k = \frac{f_1 - f_0}{T}, \tag{2}$$

where f_0 is the beginning frequency, f_1 is the end frequency, and T is the time to sweep from f_0 to f_1.

2.2 Transmission Scheme

We propose a novel transmission scheme to expand the measurable positioning area for acoustic localization. There are two main reasons why positioning cannot be performed over a wide area. First, in some scenarios, the range between a receiving device and the speaker required for positioning is quite long. It is commonly known that signal attenuation is proportional to the second power of distance. It is also known that the higher the frequency, the more it is affected by attenuation. For these reasons, these weak signals easily get contaminated by other signals and noise. Second, the receiver and speaker do not always face each other head-on. This indicates that the directivity of the speaker or the receiver or both will be poor in many environments. In particular, microphones embedded in COTS devices are greatly affected by their frequency responses because of the low performance of their electronic components.

We investigated the effect of various arrival angles on the frequency response between 7 kHz and 23 kHz using a COTS device. Figure 1 shows the frequency response of the Samsung Galaxy S10 Plus, which we use as the main device in this paper. The experiment was conducted every 15°, with 0° representing facing

each other head-on. This result clearly shows the degradation effect of the arrival angle, especially around 20 kHz. Note that the distance was kept constant in each measurement, and the results would be worse if distance attenuation was added. For these reasons, it is difficult to receive the signals when the arrival angle is large.

We utilize a transmission scheme that combines TDMA and FDMA to address the limitation. TDMA-based positioning systems have the advantage of saving the frequency band, because they schedule transmission signals from speakers to avoid band collisions. On the flip side, the positioning update rate is slow because the device needs time to wait to receive signals for positioning. Only a signal from one node can be received in each time slot when the TDMA transmission scheme is adopted. This is a major disadvantage when positioning a moving target because the position where the signal is received is different each time.

FDMA-based positioning systems, on the other hand, have the advantage of positioning update rate, because they transmit different frequency band signals at the same time to prevent interference with each other. However, because different bands are transmitted at the same time, the total band inevitably becomes wide. Even though FDMA-based systems require a wide frequency band in total, they have been adopted for acoustic-based localization systems considering the benefit of positioning update rate.

In our proposed approach, we design each speaker to transmit predefined different band chirps in a predefined sequence, as shown in Fig. 2. For example, assuming there are three speakers and performing 2-D positioning, there are four time slots per cycle for each speaker. In this case, there are three bands of chirps: high, middle, and low, and the bands do not overlap. The first time slot indicates the beginning of transmission per cycle and all speakers do not send anything. In the last three time slots, each speaker transmits a chirp that does not interfere with other signals. Knowing the transmission schedule of the speakers, the receiving device can identify each speaker.

For large-area positioning, we leverage transmitting multiple frequency bands of chirps from each speaker to suppress the effects of frequency response. For example, in an environment where positioning is severe, it is difficult to receive signals during a single time slot because of either or both attenuation and frequency response. However, by occupying multiple time slots to transmit multiple frequency band chirps, we can suppress the cause of the positioning failure.

Another major benefit of our system is that the positioning update rate is faster than TDMA-based systems as long as the environment is capable of receiving signals. This is because the signal can be detected from speakers required for positioning during a single time slot, while TDMA-based systems require multiple time slots.

2.3 Received Signal

Matched filtering (MF) is a pulse compression technique widely used in radar systems for detecting signals. In MF, a received signal is convolved with the reference signal. Let s_a denotes analytic version of the transmitted signal expressed as follows.

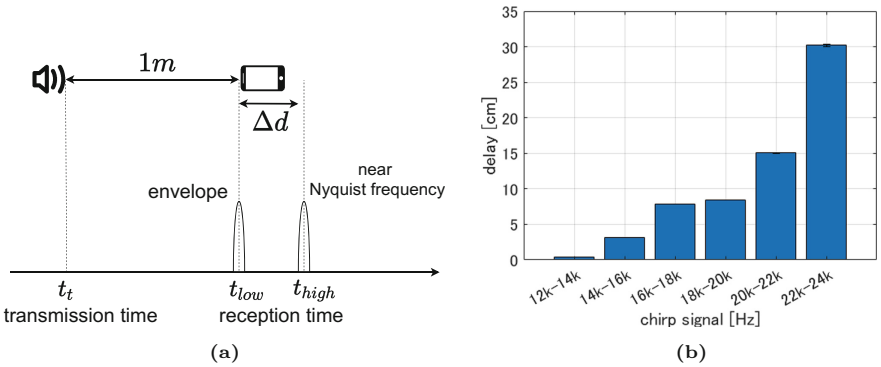


Fig. 3. (a) The reception time varies depending on the receiving frequency even though the distance is constant, (b) Ranging error caused by delay

$$s_a(t) = \exp\{j2\pi(f_0 t + \frac{k}{2}t^2)\}, \tag{3}$$

The MF output $R(\tau)$ is calculated as:

$$R(\tau) = \int_0^T s_a^*(t) r(t+\tau) dt, \tag{4}$$

where $r(\tau)$ is the received signal and $s_a^*(t)$ is the reference signal which is conjugated. We can also obtain an envelope $E(\tau)$ by calculating the magnitude of $R(\tau)$.

2.4 Group Delay

Because of the uniqueness of our transmission method, we noticed that the reception time varies depending on the frequency shown in Fig. 3a. For example, let t_t is the transmission time, t_{low} and t_{high} are the reception times of low and high frequency signals respectively, and Δd is the delay. Assuming the true distance between the speaker and the device is 1m, we can measure the range calculating $c \cdot (t_{low} - t_t)$ and $c \cdot (t_{high} - t_t)$ where c is the speed of sound. Even if the distance is the same, we observed that the reception time of the high frequency t_{high} is shifted, resulting in a longer estimated distance. Typically, anti-aliasing filter (AAF) is applied to limit near Nyquist frequency at the input of an analog-to-digital converter. This time delay is longer when recording with a COTS device that has an AAF with low performance. This time delay is called group delay [19]. The delay can be considered the propagation time delay of the envelope as it passes through a digital filter.

We conducted a preliminary experiment to investigate the delay. In the experiment, we first transmit 10–12 kHz up-chirp that is not affected by the AAF. Then, after the given time interval, we transmit a signal to be checked. A smartphone as a receiver is located 1 m away from the speaker. We compute MF to

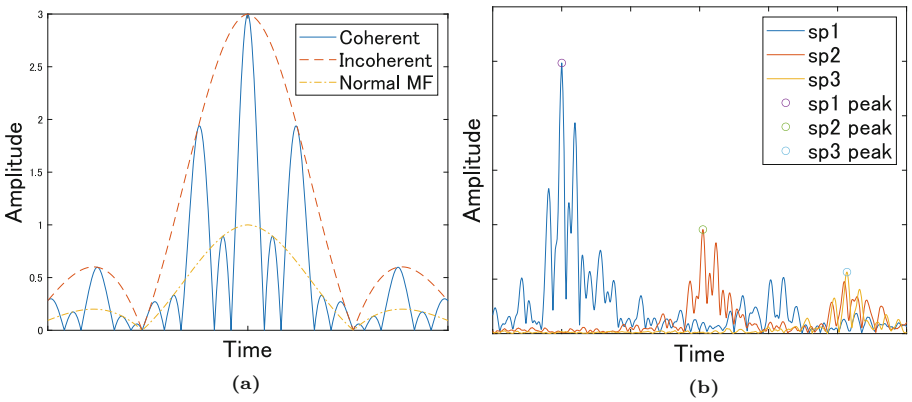


Fig. 4. (a) Simulation of the averaging techniques, (b) Envelope of actual received signal after coherent averaging

obtain the highest peak as the reception time. The time interval between the peak of the first transmitted signal and the peak of the signal being investigated should be the same as the given time interval if the group delay does not affect the signal. In this experiment, we used Samsung S10 plus, and 200 measurements were made for each signal. The results are shown in Fig. 3b. Note that the delay may vary depending on the COTS device. The results show that the delay increases as the frequency approaches the Nyquist rate. We also computed the variance of the delay, but they were all negligibly small. Therefore, the average value was used as the delay to be compensated in this paper.

2.5 Coherent Averaging

Because a speaker transmits multiple bands of signals in a single cycle (including multiple time slots), we take advantage of the combination of the received signals to improve the range resolution. In our approach, we use a technique of coherent averaging that improves SNR by collecting multiple signals and averaging them in time phase [19]. We first collect the signals in a single cycle. Second, MF is performed for each time slot using the reference signal in the same sequence as the signals transmitted by the speaker. Finally, MF outputs are added up and obtain the envelope. This implies that the smaller the phase difference between each MF output, the higher the peak will be. Near the mainlobe, the phase of the MF output is close to zero, and thus the magnitude is larger. By contrast, when the phases are not close to each other, they cancel each other out or weaken each other. Note that the envelope from other speakers can be created by changing the sequence of the reference signals.

We also compare the performance of coherent and incoherent averaging. The process of incoherent averaging averages in the envelope, while coherent averages before making the envelope. In the simulation Fig. 4a, the envelope of coherent averaging E_{pre} contains the highest and narrow mainlobe, and large sidelobes at

both sides of the mainlobe. The narrower the mainlobe, the smaller the variance in reception time becomes due to the stabilization of the peak. In contrast, the variance of reception time for incoherent averaging is expected to be larger because the mainlobe is wider.

Figure 4b is an actual received envelope after coherent averaging. As shown in the simulation, large sidelobes appear on both sides of the mainlobe. It can suppress the sidelobes but requires a wider bandwidth. We set each chirp's bandwidth considering the trade-off in this paper.

Envelopes of coherent $E^i_{pre}[n]$ and incoherent averaging $E^i_{post}[n]$ from the i-th speaker are expressed as follows:

$$E^i_{pre}[n] = \| \sum_{l=2}^{M} R^i_l[N \cdot (l-1) + n - \Delta d^i_l] \| \quad 0 \leq n \leq N, \tag{5}$$

$$E^i_{post}[n] = \sum_{l=2}^{M} \| R^i_l[N \cdot (l-1) + n - \Delta d^i_l] \| \quad 0 \leq n \leq N, \tag{6}$$

where M is total number of time slots, R^i_l is the output of MF transmitted by speaker i in the l-th time slot, N is the total number of samples in a time slot, and Δd^i_l is the delay in samples for the calibration that depends on the transmitted signal (see Sect. 2.4). These equations show that the coherent averaging maintains the phase information after addition, while incoherent averaging loses it before. Near the mainlobe, the phase of each signal is aligned, resulting in high amplitudes, but when the phase is different, the amplitudes cancel each other out. The reception sample of speaker i is estimated as follows:

$$n^i_{pre} = \underset{n}{\text{argmax}}\ E^i_{pre}[n], \tag{7}$$

$$n^i_{post} = \underset{n}{\text{argmax}}\ E^i_{post}[n], \tag{8}$$

where the maximum peak is selected as the reception time transmitted from speaker i. The reception sample n is easily converted to the receiving time by considering the sampling rate and the speed of sound.

2.6 TDoA

To avoid requiring the mobile device to synchronize with the speaker, we use a TDoA technique to estimate the position. In TDoA, estimating the 2-D coordinates requires at least three speakers. Typically, N-dimensional position estimation requires at least N + 1 nodes. Assuming that the 2-D coordinates of each speaker are known, we can set up the following system of equations:

$$\sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_2 - x)^2 + (y_2 - y)^2} = c \cdot (\tau_1 - \tau_2), \tag{9}$$

$$\sqrt{(x_1 - x)^2 + (y_1 - y)^2} - \sqrt{(x_3 - x)^2 + (y_3 - y)^2} = c \cdot (\tau_1 - \tau_3), \tag{10}$$

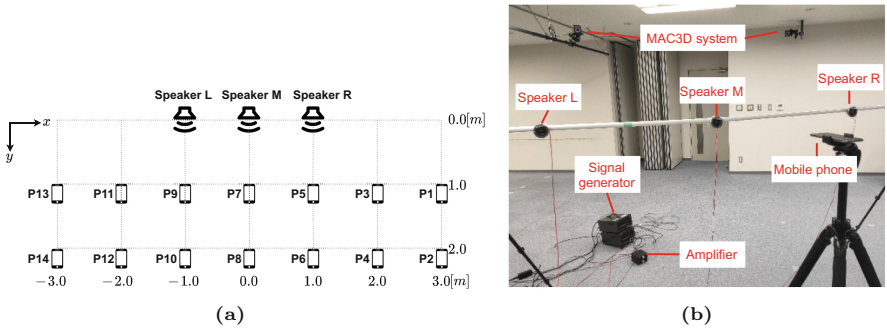


Fig. 5. Experimental environment: (a) speakers and measurement positions, (b) experiment scenario

where (x_1, y_1), (x_2, y_2), and (x_3, y_3) denote x, y coordinate of each speaker, (x, y) is the estimated position of the device, and c is the speed of sound. τ_1, τ_2, and τ_3 are reception times from each speaker. With three speakers, we can construct the presented system of two equations with two unknowns (x and y). We use the following formula to compute the speed of sound c in this paper:

$$c = 331.3 + 0.606t, \tag{11}$$

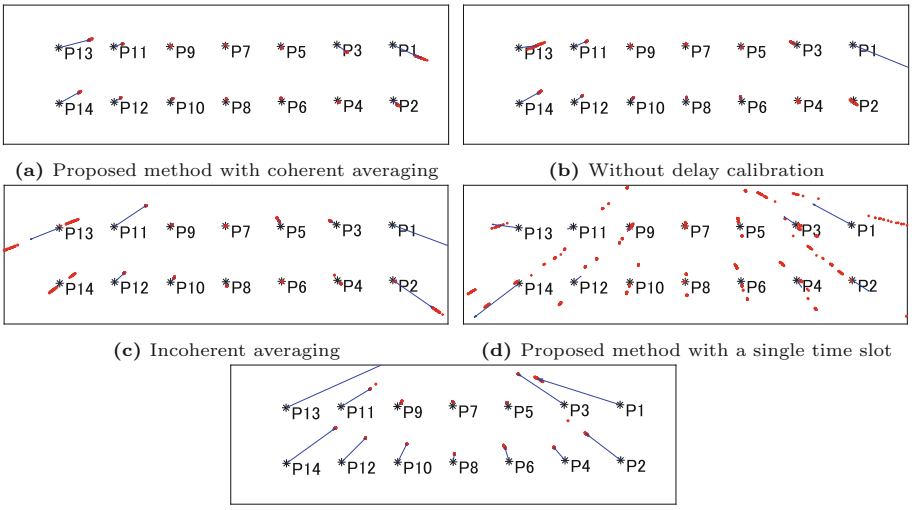
where t is the air temperature in degrees Celsius.

3 Experimental Evaluation

3.1 Experimental Setup

To evaluate our proposed method, we experimented with three speakers performing two-dimensional positioning estimation using TDoA. We used a Fostex FT200D as the speaker, an NF WF1948 as the signal generator, and a Fostex AP20d as the amplifier. We located three speakers such that they were at the same height and conducted each positioning experiment at 14 positions shown in Fig. 5a. The experimental setup can be seen in Fig. 5b. We located each speaker and smartphone facing straight ahead horizontally. We used a Samsung Galaxy S10 Plus for a COTS device. The smartphone was held by a tripod, and the height was set to be the same as the speakers to estimate 2-D coordinates. We measured the position of the speakers and smartphone as the true value using the Cortex MAC3D system, an accurate motion capture system. The temperature in the room was 20.9°C. We recorded the audio data using a recording app and set 48 kHz as the sampling rate. In the experiments, we recorded the audio data via the built-in top microphone. The recorded experimental data were applied to our proposed method offline and evaluated. Received signals are upsampled by a factor of 16 to interpolate between samples.

When setting parameters, it is important to consider trade-offs such as bandwidth and signal length. In this paper, all signal lengths were set to 10 ms, and



(a) Proposed method with coherent averaging

(b) Without delay calibration

(c) Incoherent averaging

(d) Proposed method with a single time slot

(e) FDMA-based positioning

Fig. 6. 100 estimated positions (red points) and mean positions (blue arrows) for each method: (a) Proposed method with coherent averaging, (b) Without delay calibration, (c) Incoherent averaging, (d) Proposed method with a single time slot, (e) FDMA-based positioning

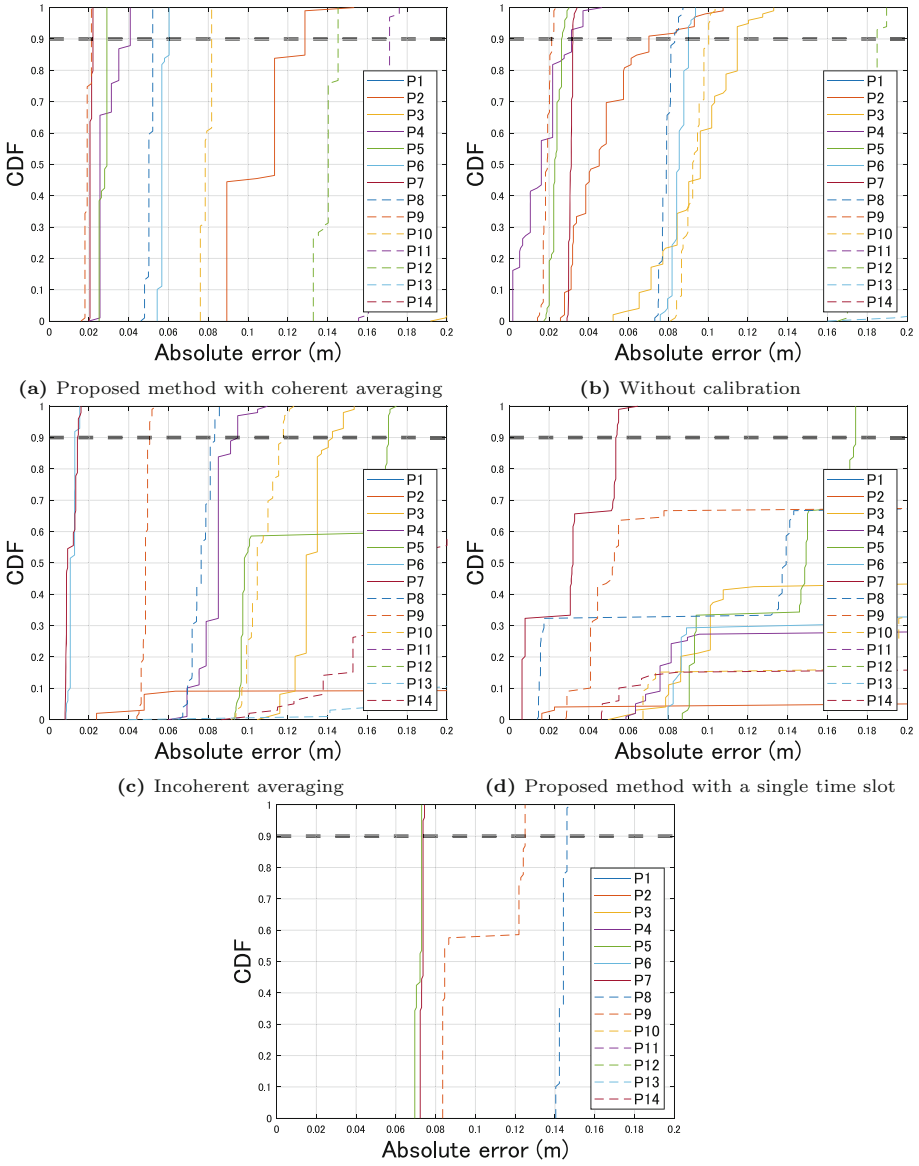
all signal bandwidths were set to 2 kHz. Aiming to avoid interference, our guard bandwidth between transmission signals was set to 2 kHz. In the experiment, we define 20–22 kHz up-chirp as high chirp, 16–18 kHz up-chirp as mid chirp, and 12–14 kHz up-chirp as low chirp. There are four time slots in total per cycle (500 ms) shown in Fig. 2, and the length of each time slot is set to 125 ms.

There were 100 measurements made in each position and evaluated. Figure 6 shows positions estimated by each experiment. Figure 7 shows the cumulative distribution function (CDF) for the measurements. Note that errors larger than 0.20 m were not included in the figure. We also defined that positioning is failed if the error was greater than 1 m or if TDoA had no real solution.

3.2 Our Proposed Approach

To investigate the positioning performance, experiments were conducted in which we applied our proposed transmission scheme. The first time slot indicates the beginning of transmission, and all speakers do not send anything. In the last three time slots, each speaker transmits a chirp that does not interfere with others. Each speaker transmitted the following predefined sequence per cycle:

- Speaker R: blank ⟶ high ⟶ mid ⟶ low
- Speaker M: blank ⟶ mid ⟶ low ⟶ high
- Speaker L: blank ⟶ low ⟶ high ⟶ mid



(a) Proposed method with coherent averaging

(b) Without calibration

(c) Incoherent averaging

(d) Proposed method with a single time slot

(e) FDMA-based positioning

Fig. 7. Cumulative error function for each positioning: (a) Proposed method with coherent averaging, (b) Without delay calibration, (c) Incoherent averaging, (d) Proposed method with a single time slot, (e) FDMA-based positioning

We applied coherent averaging to the received signals and found the maximum peaks to estimate the position per cycle. Figure 6a shows the result visually,

and Fig. 7a represents the CDF for positioning. The maximum and minimum 90th-percentile errors were 46.26 cm at P13 and 0.65 cm at P5. The maximum and minimum standard deviations were 3.13 cm at P1 and 0.06 cm at P7.

3.3 Effect of Signal Delay

We also conducted some other experiments with the same data to compare the performance. First, we conducted coherent averaging without delay calibration to identify the effect of signal delay. Figure 6b shows the result visually, and Fig. 7b represents the CDF for positioning. The positioning did not succeed every time at P1. Excluding P1, the maximum and minimum 90th-percentile errors were 46.33 cm at P14 and 2.12 cm at P4, respectively. The maximum and minimum standard deviations were 6.47 cm at P1 and 0.09 cm at P7, respectively.

3.4 Incoherent vs. Coherent

We conducted incoherent averaging applying to the same data aiming to determine the variance. Figure 6c shows the result visually, and Fig. 7c represents the CDF for positioning. Similarly, excluding P1 and P13, the maximum and minimum 90th-percentile errors were 69.75 cm at P11 and 0.89 cm at P6, respectively. The maximum and minimum standard deviations were 12.81 cm at P1 and 0.19 cm at P6, respectively.

3.5 Single Time Slot vs. Three Time Slots

We experimented without any averaging techniques to estimate the position, that is, positioning during each time slot instead of three time slots. Figure 6d shows the result visually, and Fig. 7d represents the error for positioning. Similarly, excluding P1, P2, P3, P4, P11, P12, P13, and P14, the maximum and minimum 90th-percentile errors were 34.61 cm at P10 and 5.36 cm at P7, respectively. The maximum and minimum standard deviations were 12.33 cm at P9 and 1.93 cm at P7, respectively.

3.6 Comparison with FDMA-Based Positioning

To compare the performance with the conventional method, we experimented with an FDMA-based positioning system. In the experiment, Speaker R transmitted the high chirp, Speaker M transmitted the mid chirp, and Speaker L transmitted the low chirp continuously in a loop. To make the experiment fair, each speaker transmitted three times per cycle, and coherent averaging is applied to the received signals to estimate the position. Figure 6e shows the result visually, and Fig. 7e represents the CDF for positioning. We could not estimate the position at P1, P2, P3, P13, and P14. Excluding these positions, the maximum and minimum 90th-percentile errors were 58.21 cm at P12 and 6.65 cm at P7, respectively. The maximum and minimum standard deviations were 2.18 cm at P6 and 0.07 cm at P7, respectively.

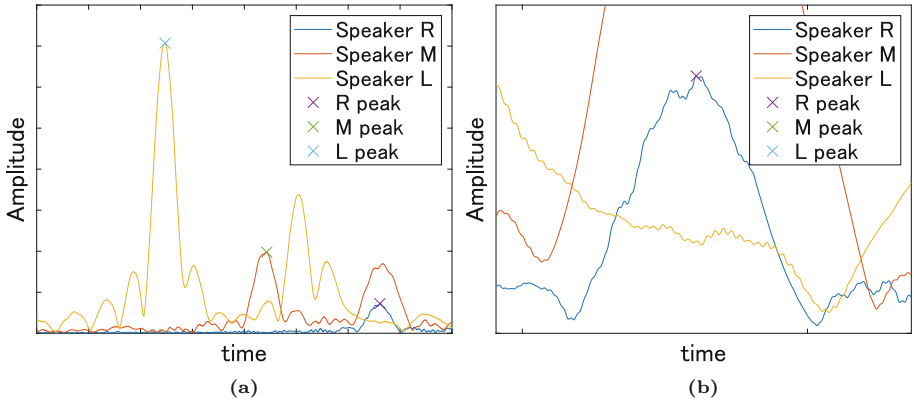


Fig. 8. (a) Envelope at P13 with FDMA positioning, (b) Enlarged view of the figure on the left

4 Discussion

Measurements at the edge positions such as P1, P2, P13, and P14 where the arrival angle from the speaker is severe make position estimation more difficult. Especially in the time slot where a high chirp is transmitted, it becomes even more difficult because of the frequency response shown in Fig. 1. However, our proposed method can mitigate the effect by using multiple frequency bands. Figure 8a is an envelope of the received signal at P13 performing FDMA-based positioning. We found that the signal from Speaker R, which is farthest from the device and with a severe arrival angle, is contaminated by the noise shown in Fig. 8b. This indicates that the peak is unstable and consequently estimates the wrong position. A similar waveform was also observed at P1, P2, P3, P13, and P14. For these reasons, the positioning error is large compared with our proposed approach.

We compared the performance without delay calibration. The results were nearly similar, but the variance is smaller for the proposed method. We believe the reason is that without delay correction, the signals do not add up properly near the mainlobe and become unstable.

We also compared the performance with positioning during a single time slot. Because positioning occurs during a single time slot, it is difficult to estimate the position over a large area. However, if the device is in an area where it can receive a sufficient signal, it is possible to estimate its position accurately. In the limited area, the positioning update rate is faster as long as the environment is conducive to signal reception. We can take advantage of TDMA and FDMA to estimate the accurate position using single time slot in the middle area where the error is little and using three time slots in the area where the positioning is difficult. However, the results show that the area is quite limited to the central area such as P5, P7, P8, and P9 where it is not easily affected by directivity.

It is difficult to estimate the position of a moving target because our proposed approach requires three time slots to perform averaging. In coherent averaging, signals are accumulated using phase information, but they are not added up properly if reception time varies in each time slot.

5 Related Work

5.1 Indoor Positioning Technologies for Smartphones

It is desirable for an indoor positioning system to be able to estimate a user's position using only a smartphone and without additional hardware. Therefore, many indoor positioning studies have been proposed for smartphones. We can summarize smartphone-based indoor positioning in terms of three categories: inertial-sensor-based tracking, radio-signal-based localization, and acoustic-signal-based positioning. The pedestrian dead reckoning (PDR) approach, which uses inertial sensors, is very useful because it is possible to estimate a relative position using only a smartphone [12]. However, it is necessary to combine it with other positioning methods because an initial smartphone position has to be acquired and positioning errors accumulate while walking [18]. In radio-signal-based localization, UWB based systems have proven to provide 10–20 cm accuracy [26], however, most of the current devices do not have UWB chip. For a smartphone system, an RSSI-based position-estimation method using Wi-Fi [2] or BLE [7] has become widely used. Furthermore, IEEE 802.11-2016 now includes a Wi-Fi Fine Time Measurement (FTM) protocol, and thus more robust approaches for Wi-Fi localization are proposed [24]. However, these methods have a positioning error of around 1 m and low accuracy. By contrast, an acoustic signal-based positioning method has a positioning accuracy as good as centimeter level. The acoustic signal enables high accuracy estimation because its propagation time is slower than a radio signal.

5.2 Acoustic Signal-Based Positioning Systems

Acoustic signals are suitable for device positioning in terms of accuracy, and many studies have been proposed.

Examples of such systems include Active Bats [10], Cricket [22], and DOLPHIN [8], which are ultrasonic 3-D positioning systems using the Time of Arrival (ToA). They can achieve highly accurate positioning. Moreover, Cricket [22] has been extended to Cricket Compass [23], which can measure the yaw angle of the target by Angle of Arrival (AoA).

It is also difficult for commercial mobile devices, including smartphones, to receive ultrasonic signals and provide high-speed and precise time synchronization between transceivers, as is required for ToA calculation. Only one example of such time synchronization has been proposed, which uses a smartphone, camera, and lighting [1]. Therefore, alternative TDoA-based methods, which do not

require time synchronization, are more widely used rather than ToA-based systems. Examples of methods that use TDoA include ASSIST [11], ALPS [16], and Sonoloc [6].

FDMA-based transmission schemes are often used for basic acoustic-based positioning systems for improving the positioning update rate. In [17], the authors try to identify speakers by modulating chirp rate adaptation to save the frequency band. Similary in [13], the authors design optimal waveforms to mitigate the effect of multiple-access interference. A FDMA-plus-TDMA-based system is also used in [5]. They propose a unique transmission scheme that saves the frequency band and improves the update rate by shifting the transmission time slightly.

There are many other methods for acoustic-based positioning [3,9,20,21]. However, to the authors' knowledge, no previous works have been researched for expanding the positioning area.

6 Conclusion

In this paper, we propose a novel transmission scheme to expand the positioning area. In the transmission scheme, FDMA and TDMA techniques are combined to compensate for the disadvantages. Outputs of MF are then added up by applying the coherent averaging to improve the range resolution and SNR. Finally, the maximum peaks are selected as the reception times to compute TDoA. We also refer to the receiving delay that is affected by AAF. We conduct a preliminary experiment to calibrate the received signals in this paper. We installed three speakers in a room and conducted positioning experiments at 14 measurement positions. The result demonstrates high-accuracy positioning and small variance for all positions, confirming the robustness of our system and enabling larger-area positioning.

Acknowledgement. This research was supported by JSPS Kakenhi Grant Numbers 19H04222 and 20K21781.

References

1. Akiyama, T., Sugimoto, M., Hashizume, H.: Time-of-arrival-based indoor smartphone localization using light-synchronized acoustic waves. IEICE Trans. Fundam. Electron. Commun. Comput. Sci. **100**(9), 2001–2012 (2017)
2. Bahl, P., Padmanabhan, V.N.: Radar: an in-building RF-based user location and tracking system. In: Proceedings IEEE INFOCOM 2000. Conference on Computer Communications. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies, vol. 2, pp. 775–784. IEEE (2000)
3. Cai, C., Zheng, R., Li, J., Zhu, L., Pu, H., Hu, M.: Asynchronous acoustic localization and tracking for mobile targets. IEEE Internet Things J. **7**(2), 830–845 (2019)
4. Carlyn, M., et al.: Effects of age, season, gender and urban-rural status on time-activity: Canadian human activity pattern survey 2 (chaps 2). Int. J. Environ. Res. Public Health **2**(11), 2108–2124 (2014)

5. Chen, X., Chen, Y., Cao, S., Zhang, L., Zhang, X., Chen, X.: Acoustic indoor localization system integrating TDMA+FDMA transmission scheme and positioning correction technique. Sensors **19**(10), 2353 (2019)
6. Erdélyi, V., Le, T.K., Bhattacharjee, B., Druschel, P., Ono, N.: Sonoloc: scalable positioning of commodity mobile devices. In: Proceedings of the 16th Annual International Conference on Mobile Systems, Applications, and Services, pp. 136–149 (2018)
7. Faragher, R., Harle, R.: Location fingerprinting with Bluetooth low energy beacons. IEEE J. Sel. Areas Commun. **33**(11), 2418–2428 (2015)
8. Fukuju, Y., Minami, M., Morikawa, H., Aoyama, T.: DOLPHIN: an autonomous indoor positioning system in ubiquitous computing environment. In: Proceedings IEEE Workshop on Software Technologies for Future Embedded Systems, WSTFES 2003, pp. 53–56. IEEE (2003)
9. Ge, L., Zhang, Q., Zhang, J., Huang, Q.: Acoustic strength-based motion tracking. Proce. ACM Interact. Mob. Wearable Ubiquit. Technol. **4**(4), 1–19 (2020)
10. Harter, A., Hopper, A., Steggles, P., Ward, A., Webster, P.: The anatomy of a context-aware application. Wireless Netw. **8**(2), 187–197 (2002)
11. Höflinger, F., et al.: Acoustic self-calibrating system for indoor smartphone tracking (assist). In: 2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN), pp. 1–9. IEEE (2012)
12. Kang, W., Han, Y.: SmartPDR: smartphone-based pedestrian dead reckoning for indoor localization. IEEE Sens. J. **15**(5), 2906–2916 (2014)
13. Khyam, M.O., et al.: Simultaneous excitation systems for ultrasonic indoor positioning. IEEE Sens. J. **20**(22), 13716–13725 (2020)
14. Krishnan, S., Sharma, P., Guoping, Z., Woon, O.H.: A UWB based localization system for indoor robot navigation. In: 2007 IEEE International Conference on Ultra-Wideband, pp. 77–82. IEEE (2007)
15. Lanjudkar, P.: Indoor positioning and indoor navigation (IPIN) Market accessed May 23 2021 (2018). https://www.alliedmarketresearch.com/indoor-positioning-and-indoor-navigation-ipin-market
16. Lazik, P., Rajagopal, N., Shih, O., Sinopoli, B., Rowe, A.: ALPS: a Bluetooth and ultrasound platform for mapping and localization. In: Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems, pp. 73–84 (2015)
17. Lazik, P., Rowe, A.: Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In: SenSys 2012 - Proceedings of the 10th ACM Conference on Embedded Networked Sensor Systems, pp. 391–392 (2012)
18. Liu, T., Niu, X., Kuang, J., Cao, S., Zhang, L., Chen, X.: Doppler shift mitigation in acoustic positioning based on pedestrian dead reckoning for smartphone. IEEE Trans. Instrum. Meas. **70**, 1–11 (2020)
19. Lyons, R.G.: Understanding Digital Signal Processing, 3rd edn. Prentice Hall, Upper Saddle River (2011)
20. Murakami, H., Suzaki, T., Nakamura, M., Hashizume, H., Sugimoto, M.: Five degrees-of-freedom pose-estimation method for smartphones using a single acoustic anchor. IEEE Sens. J. **21**, 8030–8044 (2020)
21. Nakamura, M., Hashizume, H., Sugimoto, M.: Simultaneous localization and communication methods using short-time and narrow-band acoustic signals. Sensor Device Technologies and Applications SENSORDEVICES 2020 (2020)
22. Priyantha, N.B., Chakraborty, A., Balakrishnan, H.: The cricket location-support system. In: Proceedings of the 6th Annual International Conference on Mobile Computing and Networking, pp. 32–43 (2000)

23. Priyantha, N.B., Miu, A.K., Balakrishnan, H., Teller, S.: The cricket compass for context-aware mobile applications. In: Proceedings of the 7th Annual International Conference on Mobile Computing and Networking, pp. 1–14 (2001)
24. Xu, S., Chen, R., Yu, Y., Guo, G., Huang, L.: Locating smartphones indoors using built-in sensors and Wi-Fi ranging with an enhanced particle filter. IEEE Access **7**, 95140–95153 (2019)
25. Zafari, F., Gkelias, A., Leung, K.K.: A survey of indoor localization systems and technologies. IEEE Commun. Surv. Tutorials **21**(3), 2568–2599 (2019)
26. Zafari, F., Papapanagiotou, I., Christidis, K.: Microlocation for internet-of-things-equipped smart buildings. IEEE Internet Things J. **3**(1), 96–112 (2015)