Time-of-Arrival Based Acoustic Localization Technique Using Visible Light Communication for Smartphones

AKIYAMA Takayuki

Doctor of Philosophy

Department of Informatics
School of Multidisciplinary Sciences
SOKENDAI (The Graduate University for Advanced Studies)

Time-of-Arrival Based Acoustic Localization Technique Using Visible Light Communication for Smartphones

AKIYAMA Takayuki

A dissertation submitted to the Department of Informatics
the School of Multidisciplinary Sciences
in partial fulfillment of the requirements for the degree of

Doctor of Philosophy

SOKENDAI
The Graduate University for Advanced Studies

September 2017

Abstract

We describe a precise and autonomous indoor localization technique for smartphones. Almost everyone owns a smartphone nowadays. As a person always carries it with him/her, it can be a good tool for locating a position, or localization. Localization is a key technology for location-aware systems, which provide us information according to the location.

We can use a Global Navigation Satellite System (GNSS) for outdoor localization. However, GNSS's radio frequency signals hardly penetrate into rooms. Therefore, another means is required for indoor localization. Smartphones have many kinds of sensors, such as accelerometers, gyroscopes, and geomagnetic sensors. A Wi-Fi transceiver, a Bluetooth transceiver, video cameras, and microphones are also sensors. We can use these sensors to achieve indoor localization.

In general, indoor localization requires more precise information than that of outdoors. GNSS localization has an accuracy of few meters. It is enough for outdoor localization applications, such as car navigation systems. However, it is coarse for indoor localization even if GNSS could be available indoors. For example, it is difficult to distinguish adjacent doors in a room with this accuracy.

Our goal is to attain accuracy and precision of centimeters. Then, we can realize precise indoor navigation that can alert users to steps or stairs. We can guide users to the targeted shelves of goods in a large warehouse. Tracking a smartphone in one's hand, we are also able to make a machine interface such as motion control.

To accomplish this goal, we developed an acoustic localization system using smartphones. A transmitter consists of loudspeakers emits short acoustic waves, and a smartphone's microphone catches them. As rooms usually have audio equipment, we can make use of it for the sender node. The frequencies of the acoustic waves are in a higher range that is almost inaudible for human ear. However, they can be handled by commercial off-the-shelf loudspeakers and smartphones. The localization system is based on the time of flight of each acoustic wave. We conducted experiments in a 2D arrangement and confirmed the feasibility of this method.

Although the feasibility of the acoustic localization was established, the precision was not satisfied for our requirements. Especially in the sender-receiver direction, the precision was about 150 mm to 200 mm, although the precision in its perpendicular direction was around 10 mm. This uneven distribution was caused by the localization scheme of time difference of arrival (TDoA), which is deployed when the sender and the receiver are not synchronized. Employing time synchronization, we can improve the localization precision with using time of arrival (ToA) scheme.

To realize ToA localization, we devised a time synchronization method using LED light. Almost every room has lighting facility. These days, LED light is commonly used because of its low energy consumption and long-term performance. Now, we modulate LED light. With using a switching circuit, LED is easily modulated at even high frequency that we cannot detect. Smartphone's video camera captures modulated LED light that is a part of the sender node, then the smartphone extracts time information from the captured video images of the modulated

LED light. In this situation, we can assume that the time synchronization with the transmitter is established for the smartphone. 3D localization experiments were conducted for both ToA and TDoA. ToA showed better performance than TDoA, and the precision was not more than 10 mm.

Finally, to realize the autonomous localization, we deploy visible light communication (VLC) for time synchronization. VLC conveys information on modulated light. VLC's demodulation process includes clock detection, which is inherently equivalent to time synchronization. Therefore, modulation for only time synchronization is wasteful. We can use modulation for also data transmission.

We are able to transmit the data needed for localization, such as coordinates of the sender beacons and the frequencies of acoustic waves using VLC. In this case, we do not have to preset or renew those data in a localization application. In addition, we need not inquire them to external databases that are generally implemented as cloud services.

Experimental results of a 3D localization application using VLC time synchronization showed the precision of around 100 mm to 200 mm, which could be reduced to 10 mm to 20 mm by compensating the frequency difference between the transmitter and the smartphone. It was rather worse than the case of time synchronization using dedicated LED modulation pattern. However, the precision is good enough for practical applications.Consequently, we have attained a precise and autonomous indoor localization technique.

Contents

1. Introduction

1.1. Background and Motivation

Smartphones have prevailed into people. Maps are popular smartphone applications because they indicate where you are in an outdoor environment. Almost smartphones have a receiver for global navigation satellite systems (GNSSs) such as the global positioning system (GPS). Smartphone navigation applications using GNSSs are now widely used. Previously, car navigation systems that require dedicated GNSS receivers were the only application as personal navigation systems. Today, using smartphones' GNSS function, one can determine its position on a map, and get a route to a destination outdoors. Because many smartphone map applications are free, they are taking over dedicated GPS navigation systems.

Although GNSS is useful outdoors, it is not available indoors because the GNSS's radio frequency signals do not reach rooms in buildings. Therefore, other means should be developed to locate ones' position. To address this requirement, we have devised a precise and autonomous indoor localization system for smartphones.

Indoor localization needs precision higher than that of outdoor localization. For example, in many cases, it is enough to distinguish buildings for car navigation. Indeed, GNSS localization has an accuracy of few meters. However, it is coarse for indoor localization walking around a room. At least, we need to identify corners or doors in corridors or rooms. It is desirable to alert senior citizens or people with disabilities to steps or stairs. To achieve these requirements, we require the precision of centimeters. With this precision, we can determine the shelves where small items exist in a large warehouse. In addition, tracking the smartphones' position, we can introduce a man-machine interface such as a gesture input.

Autonomous measurement is the other issue to be considered. An off-the-shelf GNSS receiver can conduct localization autonomously, i.e., a receiver can determine its position only using received signals. There is no interaction between the satellite systems and the receiver. The receiver does not need to consult centralized servers or databases.

There are some advantages of this scheme. First, the whole system is simple. The satellite system just sends signals, while receivers only accept the signals. Although the satellite system is synchronized, no interactions between the sender and the receivers exist. Nor, interactions among the receivers exist. Next, this scheme is scalable. Sender satellites can be added at an arbitrary time, and if there is enough number of satellites, an outage of a satellite has a small impact on the receivers. In addition, increasing the number of receivers does not affect the satellite system. Finally, we can protect the privacy of the users. Localization is done without announcing the presence of the receivers.

Our indoor localization system also aims at this scenario. Localization process should be done autonomously, and the interaction between the sender and the receiver should be avoided. Information needed to localize the position is to be transferred one-way from the sender to the receiver.

1.2. Overview of the Proposed System

We propose an acoustic indoor localization system for smartphones. Using time synchronization with LED light, we can implement time-of-arrival (ToA)-based localization that performs precise measurements. Because signals and information for localization are sent only from the transmitter to the receiver, the receiver can fulfill autonomous measurements.

The transmitter, the *anchor node*, has at least three loudspeakers, *beacons*, for a 3D arrangement, and emits short acoustic waves, *bursts*. The receiver, a smartphone in this case, catches them and calculates the time of flight (ToF) between the beacons and the receiver. Multiplying the speed of sound, ToF can be converted to the distance; therefore, ToF is equivalent to the distance. Then using *trilateration*, we can determine the position of the smartphone. Trilateration is a process to fix the unknown vertex by making a triangle with three edges.

The frequencies of the bursts are, for example, in a range of 13 kHz to 17 kHz, which are almost inaudible for human ear, although they can be dealt with by commercial off-the-shelf loudspeakers and smartphones. To detect the acoustic signals, we use the phase accordance method (PAM) [15]. PAM was originally developed for ultrasonic waves around 40 kHz, however, it is also applicable for audible acoustic waves. PAM uses short beat waves, of which the duration is 2 ms to 4 ms. It can precisely determine the position of the center of each wave. The resolution is finer than the sampling interval that is 48 ksps, for example. This leads to precise ToF measurements. Because the duration of each wave is short, signal detection can be done rapidly, and the update frequency of the measurement can be increased.

As mentioned above, the transmitter consists of multiple loudspeakers. To make the transmitter small-sized, the separations of the loudspeakers, or the *baselines*, should be short. The loudspeakers emit acoustic waves that have different frequencies each other. Smartphone's monaural microphone catches each wave using PAM and calculates ToF. Then, the smartphone obtains the distances between the loudspeakers and the smartphone. Because the baselines are short, acoustic waves, which are emitted from the loudspeakers at the same time, are received almost simultaneously.

There are two major localization schemes that use ToF measurements, ToA and time difference of arrival (TDoA). If the smartphone gets the emission time of a sound wave, it can calculate the ToF of the wave because it knows the reception time of the wave. Now that we know the absolute distances to the loudspeakers from the smartphone, we can determine the position of the smartphone. This is the ToA localization. ToA needs time synchronization between the transmitter and the receiver, because the receiver should know the transmitter's sound emission time. At least three loudspeakers are needed in 3D ToA localization. In the case time synchronization is not established, we can deploy TDoA localization. The points that have the same TDoA mathematically make a hyperbola in a 2D arrangement or a hyperboloid in that of 3D, of which the foci are the positions of the loudspeakers. Getting two hyperbolae in 2D or three hyperboloids in 3D, we can determine the smartphone's position as the intersection of them. TDoA needs one more loudspeakers than ToA. ToA's localization precision is better than that of TDoA, especially when the baselines are short. We adopt ToA for our localization.

As we mentioned above, time synchronization is needed for ToA localization. We use LED light for this purpose. We modulate LED light with some pattern. Capturing it by a smartphone's video camera, we detect the time when the modulation is started. This is the situation we assume that the smartphone is synchronized with the transmitter in the study. We mount loudspeakers and an LED light onto the transmitter and emit acoustic waves and

modulated light simultaneously. Because the speed of light is much faster than that of sound, we can ignore the propagation time of light in acoustic localization. Therefore, the receiver can obtain the emission time of the acoustic waves, and then can calculate the ToF for ToA localization.

We conducted 3D localization experiments for both ToA and TDoA using time synchronization. The results were as follows. In ToA, the precision was not more than 10 mm when the distance between the transmitter and the smartphone was around 2 m. Although some systematic errors were observed, they can be reduced by calibration. In contrast, in TDoA, the localization errors were much larger, and the accuracy was also worse than the case of ToA. The precision of time synchronization was equivalent to 0.17 mm of an airborne acoustic wave. ToA's better performance was based on this precise time synchronization. The availability of time synchronization and ToA localization was confirmed

Time synchronization using video images is not an easy task. The frame rate of the smartphone's video cameras is usually 30 fps–60 fps, which is too coarse for time synchronization. In other words, recognizing the switching of the LED light frame by frame is not sufficient for time synchronization to perform precise localization. To overcome this issue, we exploit the image sensor's *rolling shutter* effect to extract precise time information from an image. Smartphones usually have complementary metal-oxide-semiconductor (CMOS) image sensors for video cameras. CMOS image sensors have rolling shutters, which scan a video image sequentially line by line. Because the acquired image is not the exact snapshot at a certain time, the captured picture is distorted when the object is moving. The rolling shutter effect is usually considered to be avoided. However, because each line in an image has a different exposure time slot, we can analyze the lines as a time series. Then, we improve the time resolution to at least one thousand times the original frame rate since an image usually has around 1000 lines or more. This is the principle of the time synchronization mechanism.

We use LED modulation for time synchronization. However, it is much more used for VLC. Indeed, the rolling shutter effect has been utilized for VLC. VLC's demodulation process inherently includes clock detection, which is equivalent to time synchronization in this study. Now, we deploy VLC for time synchronization in ToA localization.

Using VLC, we send information needed for localization such as coordinates of the loudspeakers and the frequencies of sound waves. These data could be obtained by inquiring from external systems, as preset data in an application, or through other communication methods. Inquiring is not suitable for autonomous localization, because we depend on external databases and the connection to them is necessary. Presetting is also not preferable; maintaining all the data around the world as application's internal data is not practicable. We get information through VLC in this study. This method has an advantage that the user's privacy is protected because its presence is hidden in a localization process.

We conducted 3D ToA localization experiments using VLC. The precision was 100 mm to 200 mm, though it could be reduced to 10 mm to 20 mm. This is worse than the experimental results using a dedicated LED modulation pattern; however, it is precise enough for practical applications.

We are expecting some applications using our technique. First, it is applicable for indoor navigation. There are corners, adjacent doors, steps, and stairs in rooms and corridors. The proposed method has enough precision to guide users in these situations. The transmitter consists of off-the-shelf loudspeakers and an LED light. Existing acoustic and lighting equipment can

be diverted to the transmitter. Therefore, we can realize the localization system with minimum expense.

By tracking the motion of the smartphone, it is also applied to a user interface for a gesture-based control system. Interacting applications for multiple users can be realized. If smartphones are put side by side, we can distinguish between them. Therefore, by tiling smartphones that are independent of each other, virtual large display can be built.

1.3. Contributions

We discuss indoor localization for smartphones. It includes time synchronization for precise localization and VLC for autonomous measurements. The contributions of the dissertation are as follows.

- Acoustic ToA measurement system using time synchronization is proposed. The results showed better performance than TDoA. The error distribution is compared with an analytical calculation and the consistency of them is observed.
- A time synchronization method using a modulated LED light and a smartphone's video camera is proposed. Analyzing video images, we have achieved less than one-thousandth finer precision of video camera's frame rate. The synchronization process is done by one-way manner. We also show how to align video frames and audio ones.
- Time synchronization using VLC is proposed. Transmitting parameters needed for localization, a stand-alone localization scheme is established. Experimental results show enough precision for practical localization.

1.4. Dissertation Organization

The organization of the dissertation is as follows.

Chapter 2 overviews the methods for localization and time synchronization. It also describes the advantages of the proposed methods.

Chapter 3 shows an acoustic localization method for smartphones. We introduce the "Frequency Division Multiplexing Phase Accordance Method" to receive a couple of sound waves simultaneously for localization. Then we show a TDoA-based 2D localization experiment of which the data were recorded by a smartphone.

Chapter 4 shows a time synchronization technique using LED light for ToA localization. We conducted 3D localization experiments, and compared the performance of ToA with TDoA. The result was compared with the analytical distribution.

Chapter 5 shows a time synchronization technique using VLC. We perform 3D ToA localization with VLC time synchronization using a smartphone application. We discuss how we transmit parameters needed for localization on VLC.

Chapter 6 summarizes the dissertation and discusses future work.

2. Fundamentals and Related Work

2.1. Overview

In this chapter, we describe some methods of localization and time synchronization for smartphones. Smartphones have many kinds of sensors, such as a GNSS receiver, a Wi-Fi transceiver, a BLE transceiver, accelerometers, gyroscopes, geomagnetic field sensors, microphones, and cameras. Therefore, in accordance with variety of sensors, a number of localization and time synchronization methods exist. We will take a look at some of them.

Here, we consider the interaction between the measurement system and the smartphone. Let us think about acoustic localization using ToF. One scenario is to emit sound signals from a smartphone's loudspeaker, and then microphones in the rooms catch them. The measurement system collects signals from the microphones and calculates the smartphone's position using ToF. The result of the localization will be informed to the smartphone by some media, e.g. Wi-Fi. In this case, because the localization is processed outside the smartphone, the total computational load of the smartphone could be reduced. In addition, localization algorithms might be upgraded in the system. However, as the number of target smartphones increases, the measurement system needs more computational capacity. Furthermore, especially in ToF measurements, the system must distinguish each smartphone's signal. It might have a difficulty in scalability. The other scenario is the case a smartphone is the receiver. Loudspeakers in a room emit acoustic waves then the smartphone catches them and locate its position. Given that the coordinates are known, the smartphone can perform localization without interaction with the transmitting system. In this case, system's scalability problem does not come up.

We call the former case *active* localization because the smartphone itself emits acoustic waves. In contrast, we call the latter *passive* localization. We prefer to passive rather than active because of the scalability of the system. Active systems are based on two-way communication. Information is transmitted in both ways. Therefore, the number of users is restricted by the system's capacity. On the other hand, passive systems are the forms of broadcasting. Users just take sound waves broadcasted in rooms and perform localization. The number of users is not limited by the system's resources.

In addition, active localization has a disadvantage that the smartphone announces its presence to the system. In contrast, in a passive method, the smartphone itself does not emit signals for localization. Therefore, the privacy of the user is protected.

Table 2.1 summarizes the comparison between active and passive systems.

2.2. Localization Techniques

We describe localization methods available for smartphones. We discuss them from a variety of aspects. Because we pursue precise indoor localization, precision is the most important matter of us. In real applications, accuracy could be more important than precision. However, accuracy

Table 2.1. Comparison between active and passive systems.

Type	Active	Passive
Data transmission	Communication	Broadcasting
Signal direction	Two-way	One-way/None
Privacy	Weak	Protected
Application implementation	Easy	Hard
Simultaneous localization	Hard	Easy
System structure	Centralized	Distributed
System cost	High	Low

should be improved by calibration. Therefore, we mainly evaluate experimental results using precision.

2.2.1. Global Navigation Satellite System (Global Positioning System)

A Global Navigation Satellite System (GNSS) is a general term for a positioning system that uses satellites [8]. Although the name of the United States' Global Positioning System (GPS) is more popular, GPS is a kind of GNSSs. GNSSs include European Union's GALILEO, Russia's Global Navigation Satellite System (GLONASS), and China's BeiDou. Japan's Quasi-Zenith Satellite System (QZSS) is also categorized as a GNSS.

GNSS systems have scalability. In our classification, the GNSS system is the type of broadcasting, and the localization is passive. The satellite systems and the receivers are independent. Therefore, if the satellite system has enough number of satellites, increasing or decreasing number of active satellites does not affect receiver's localization performance. Even if the number of GNSS receivers increases, the satellite system need not be reinforced.

GNSS receivers cannot necessarily handle all the GNSS signals above. Old receivers might receive only GPS signals [6]. Diggelen et al. [55] illustrate the difference between GNSS systems with regard to receiver components.

As most smartphones have GNSS receivers, GNSS is widely used for outdoor localization. Its accuracy is the order of few meters, which is good enough for outdoor applications. However, the precision gets worse in urban canyon situation because of satellite invisibility and multipath [6]. Furthermore, in indoor situations where GNSS signals cannot reach, GNSS functions do not work.

2.2.2. Acoustic Waves

Characteristics of Acoustic Waves

Using acoustic waves, ToF-based localization is performed. Smartphones' loudspeakers and microphones can be used as the transmitters and the receivers, respectively. Furthermore, audio equipment in rooms can be utilized for the sound transmitter system in localization.

The speed of sound, including ultrasound, in dry air is about 340 m/s at 15 °C. Saad et al. [49] mentioned in their ultrasound localization study that this low propagation velocity leads high accuracy ToF measurements. We consider that the speed of sound is moderate for indoor

localization and its real-time software processing. Sound waves' attenuation in air and isolation by walls are advantages for indoor localization to control measuring domains.

Frequencies of sound waves should be out of the human audible range. In general, the range is assumed from 20 Hz to 20 kHz, although most adults have thresholds above about 15 kHz [41]. The audio sampling rate of smartphones is usually 44.1 ksps or 48 ksps, therefore according to the sampling theorem, the upper limit of the frequency that the original signal is to be restored is 22.1 kHz or 24 kHz. Consequently, higher frequencies ranged 15 kHz–22 kHz or 15 kHz–24 kHz could be suitable for acoustic localization. Yang et al. [59] examined the smartphones' frequency responses and then used a frequency range of 16 kHz–20 kHz for their experiments. Filonenko et al. [10] reported that smartphones' loudspeakers could generate sound signals of 17 kHz–22 kHz. Many other studies use the range of 16 kHz–20 kHz for sending or receiving [33, 18, 42].

There are some disadvantages of acoustic localization. Multipath propagation generated from room reverberation affects the performance of indoor localization, although higher frequency sound waves, which are used for acoustic localization, attenuate faster in air than lower frequency ones [20]. Noises of which the frequency ranges include the same with the measurement waves disturb the precision of the localization. It is widely known that the speed of sound is affected by the air temperature. However, it is also affected by the humidity of air [3]. Therefore, precise ToF measurements need these environmental conditions. The Doppler effect should be considered when the targets are moving.

Ultrasound has been also studied for indoor localization [56, 45, 46, 51, 40]. However, because of the upper limit of the smartphone's handling acoustic frequency, dedicated devices are needed for sending or receiving ultrasound. Therefore, we will not deal with ultrasound localization systems in detail.

Time-of-flight-based localization

We discuss two types of ToF-based localization methods; ToA-based localization and TDoA-based one. In ToA, we use absolute ToF values to localize the target. In contrast, in TDoA, we use the difference of ToF values. ToA's precision is better than TDoA's [29, 37]. In this section, we assume that there are a number of anchor nodes as transmitters and a smartphone to be located.

In ToA, to calculate ToF values, we should know the sound emission time and its reception time. When the transmitter and the receiver are time synchronized, or they have the same clock, the ToF value is calculated from the emission time by the transmitter's clock and the reception time by the receiver's clock. Because both clocks are the same, if the sound emission schedule is known, the receiver can determine the ToF. The geometrical points that have the same ToF make a circle centered at the transmitter in a 2D arrangement. If we install at least two transmitters and measure the ToF values of the sound waves emitted from the transmitters, we can locate the receiver's position at the intersection of the circles. In a 3D arrangement, the same-ToF points make a sphere, and using at least three transmitters, we can get the target's position at the intersection of the spheres.

In TDoA, we use the difference of ToF values. In this case, time synchronization between the sender and the receiver is not needed. Calculating the time difference of acoustic waves arrived at the receiver side, we can locate the position of the receiver. In a 2D arrangement, we need at least three transmitters, i.e., three arrival timestamps, to get two TDoA values. The

geometrical points that have the same TDoA make one half of a hyperbola. Receiver's position is determined as the intersection of the hyperbolic curves. In a 3D arrangement, the same-TDoA points make one half of a two-sheeted hyperboloid, and using at least four transmitters, we can get the target's position at the intersection of the curved surfaces.

Figure 2.2 summarizes the differences between ToA and TDoA. An intuitive comparison of precision is illustrated in 3.4.2. Numerical stability is discussed in 4.4.3. Considering the better precision, we aim to realize a ToA localization system.

Table 2.2. ToA and TDoA localization.

Scheme	ToA	TDoA
Precision	Better	Worse
Need synchronization	Yes	No
Number of anchors in 3D	3 or more	(3 + 1) or more
Numerical stability	Better	Worse

There have been studies on ToA-based systems, many of which use dedicated hardware for time synchronization. Harter et al. [14] used a small unit called *Bat*, which sends ultrasound pulses triggered by radio-frequency (RF) signals emitted from a base station. Receivers synchronized to the system record the ToF and then determine the Bat's position using multilateration. Accuracy at the 95th percentile was 9 cm. Minami et al. [39] proposed *DOLPHIN*, a distributed multilateration system that uses bi-directional ultrasonic transducers. The master node sends an RF message to all nodes for time synchronization. After receiving this message, a transmitter node sends an ultrasonic pulse that is picked up by receiver nodes that then calculate the distance to the transmitter. The positioning accuracy is better than 2–3 cm. *BeepBeep* [44] ranges between two mobile devices, each of which sends and receives a sound signal. By exchanging time information, the devices can estimate the two-way ToF between them. Experiments showed that the accuracy was typically better than 3 cm and that the standard deviation was less than 1 cm for most positions.

Various TDoA-based localization techniques have been proposed. For example, Höflinger et al. [17] presented a smartphone localization system to help people with visual impairments navigate. The smartphone emits high-frequency chirps that are inaudible to the human ear. The receiver system calculates the position and sends the information back to the smartphone. The positioning error is 9 cm, and the standard deviation is the same. Murata et al. [42] showed a TDoA-based smartphone localization system. The smartphone sends inaudible 17 kHz sound waves that are detected by receivers. A personal computer (PC) collects the ToF of each receiver and determines the smartphone's position.

In some studies, the meanings of ToA and TDoA are different from ours. Kim et al. [24] proposed a location tracking technique using ultrasonic signals. It takes account of ultrasonic reflection to achieve higher location accuracy. Although the paper says it is TDoA-based, we consider it ToA-based because the time difference in that paper is the ToF difference between RF and ultrasonic signals, and we can assume the RF signal is used for synchronization. In some cases, ToA measurements mean all ToF-based measurements including ToA that utilizes pseudorange measurements [36, 35, 25]. We also consider them as variations of TDoA. GNSS localization involves *ToA*-based localization using pseudoranges, which are ranged values with

the same clock offset between a satellite system and a user. This localization solves equations with four unknowns: the user's location in 3D coordinates and the clock offset. The latter is canceled by subtracting two range measurements. This method is equivalent to the TDoA localization in our study [54].

2.2.3. Image-based

Image-based localization techniques that use a smartphone camera are proposed. The position of the target is localized from the geometric arrangement of the scene. There is a study [4] which examined the AR Marker and its variations for localization. Modifying the marker patterns, they improved the localization accuracy. However, these markers are artificial and inappropriate to stick on walls or columns. Another marker-based system [50] developed wallpaper patterns that are not obtrusive but have positional information. They reported that recognition rate was highly decreased when captured with greater camera angles and distances.

A marker-less system *MoVIPS* [57] takes an image, or a low-quality video image of the surroundings, then sends them to a server. The server compares the images with a database, which have correctly located and oriented images, and then determines the location using the proportion of an obtained image to the stored one. The experimental results showed that the median position error was 0.68 m when a still image was used, and 2.85 m when a video image was used. This system needs initial data acquisition and calibration process, and storage for the database; those costs are not low.

2.2.4. Visible-light-communication-based

Image-based localization using a camera has an affinity with VLC. *Epsilon* [30] is a Received Signal Strength Indication (RSSI)-based localization using VLC. LED bulbs, which have lighting role, send information needed for localization such as the position of the bulb and the duty cycle. A receiver detects the light, measures the RSSI for each bulb, and calculates the distance. Finally, it determines its position using the information from all light sources. Experimental results showed the 90th percentile accuracy was 0.4 m–0.8 m for typical office environments. Although the receiver is a smartphone, an external light sensor board connected to the audio jack is used. This is because the sampling rate of the internal light sensor was low; the authors did not intend to use smartphone's internal camera.

Kuo et al. [26] introduced *Luxapose*, angle-of-arrival (AoA)-based localization system using VLC. VLC beacons broadcast their identities and/or coordinates. A smartphone takes pictures periodically. It examines obtained images to determine the beacons' positions and decodes VLC messages to extract beacons' identities and coordinates. Then, it locates its position by AoA localization. They demonstrated around 10 cm accuracy and 3-degree orientation error. This excellent result was obtained under controlled static conditions, in rather small region (1.0 m $\times$ 1.0 m $\times$ 2.46 m (height)), and using five LED lights though three is the minimum number of the LED lights for this situation. In the case using three LED lights, the performance was almost the same as Epsilon.

2.2.5. Received-signal-strength-indication-based

Received signal strength indication (RSSI) is mainly used for RF signals such as Wi-Fi and Bluetooth. RSSI is assumed as a monotonically decreasing function of the distance, it can be used for localization. However, because of the multipath effects of RF signals, RSSI-based indoor localization is not accurate. Therefore, RSSI fingerprint-based localization has been developed. It shows better performance than simple RSSI-based localization. Nevertheless, the drawback of this method is high cost for updating fingerprint databases and constructing RSSI radio maps [58].

RADAR [2] is a Wi-Fi RSSI-based localization system. This system based on empirical signal strength measurements and a signal propagation model. Experiment was conducted on the floor in a building that has dimension of 43.5 m × 22.5 m including more than 50 rooms. Using three Wi-Fi base stations, target laptop computer was localized. First, RSSI information for each base station was collected at 70 locations with 4 directions. Then localization process was performed. The RSSI obtained by the target laptop was matched against the collected RSSI database, and target's position was determined by triangulation. The median error was 2 m–3 m, which is about the typical office room size. Furthermore, localization that uses signal propagation model was also conducted. The median error was 4.3 m; which was rather worse than the empirical method; however, the setup costs are lower because detailed RSSI data collection is not needed.

In RF technologies, Bluetooth Low Energy (BLE) is becoming popular because of its low energy consumption. This is developed for short-range applications, and the lifetime of its device powered by a coin cell battery is more than 10 years theoretically [12]. Because BLE is short-range-oriented, we need more BLE beacons, when compared with Wi-Fi system, to localize a target in a large room. However, since BLE beacons are smaller and cheaper, it is applicable to stick many beacons in rooms. [53] demonstrated an RSSI-based 2D localization using a BLE propagation model. Experiment was conducted using three BLE beacons and a target smartphone. The results showed the accuracy of 0.4 m in an empty 4 m × 4 m room.

2.2.6. Pedestrian Dead Reckoning

Pedestrian dead reckoning (PDR) is a user tracking method using inertial sensors [13]. Localization data to be obtained are relative values, i.e., the displacements from the starting point. Furthermore, offset errors will be accumulated. Therefore, to get an absolute and accurate position, calibration is necessary at times. Seeing from another point of view, our localization system and PDR complement each other in a large indoor environment. Our system can offer precise absolute location information to PDR system. In contrast, we can use PDR to interpolate the position in the intermediate areas where acoustic waves cannot reach.

SmartPDR [21] is a PDR system that uses smartphone's accelerometer, magnetometer, and gyroscope but does not require any anchors or physical map information. Experiment was conducted in an office environment with total length of 168.55 m. The largest locating error was below 2 m and the trajectory revealed the corridors correctly.

2.2.7. Summary

Table 2.3 shows the pros and cons of various localization techniques. GNSS is not applicable in indoor environment. Acoustic systems are precise. However, if the system is active type,

i.e., it needs centralized servers, the cost is high. Image-based and VLC systems have moderate precision. Power consumption of them is high because they use video cameras for a certain period. Wi-Fi and BLE are rather coarse and they cannot be deployed in the area where RF is prohibited. However, they have some advantages. Wi-Fi signals' coverage is wide, and BLE beacons' power consumption is low. PDR has fairly good precision, however, it usually requires getting the absolute coordinates.

We employ acoustic signals for localization because we can expect precision in the order of centimeters. Especially, in passive configuration, the total system cost is not high since we can utilize audio facility in rooms. Although acoustic localization works well alone, it can be used to provide the reference position for PDR. The combination of acoustic localization and PDR could be a compromise for cost and precision.

Table 2.3. Localization media and methods.

Media/Method	GNSS	Acoustic	Image	VLC	Wi-Fi/BLE	PDR
Indoor	No	Yes	Yes	Yes	Yes	Yes
Precision	< 10 m	< 10 cm	≈ 1 m	< 1 m	1–5 m	1–2 m
Power consumption	High	Low	High	High	High/Low	High
Coverage	Global	Moderate	Narrow	Moderate	Wide/Narrow	Wide
Additional system cost	No	Moderate	Moderate	Moderate	None/Low	None
EMI-sensible situation	No	Yes	Yes	Yes	No	Yes
Absolute position	Yes	Yes	Yes	Yes	Yes	No

2.3. Time Synchronization

In this paper, synchronization means that a mobile node can estimate the time, using its own clock, when each anchor node takes an action. Anchor nodes, or beacons, are triggered internally or externally and then send signals. Beacons may or may not have their own clocks, but a smartphone has its own precise clock. If the beacon's signal emission time is detected by the smartphone's clock, we can assume that the beacon and the smartphone are synchronized.

The performance of time synchronization affects that of ToA localization. Sound waves travel 0.34 mm in 1 µs. Therefore, to achieve localization precision of centimeters, we need time synchronization that has the precision of microseconds.

2.3.1. Network Time Protocol and Precision Time Protocol

Network Time Protocol [38] is widely used to synchronize computer systems on the Internet. Giustina et al. [11] reported the performance of NTP on a Wi-Fi link, which is the connectivity smartphones usually use. Experimental results showed that the standard deviation was 1 ms, which is too coarse for ToA's time synchronization.

Precision Time Protocol (PTP) [19] offers better performance than NTP. Ferrari et al. [9] reported the performance of PTP. The standard deviation was few microseconds on a directly wired connection.

FlockLab [31] is a testbed for running multiple services simultaneously and synchronously. Servers are synchronized by NTP on Ethernet or over Wi-Fi. They say that if they add the Wi-Fi-connected server, timing error increases significantly due to variable delays via Wi-Fi.

Smartphone's Wi-Fi connection is not suitable for network-based time synchronization protocol because of the instability of the latency. Furthermore, NTP and PTP have a disadvantage that they take few hours for convergence.

2.3.2. Radio Frequency

Cricket [45] is a room-level location detection system that uses RF and ultrasound. Using the difference in RF and ultrasonic signal propagation speed, the listener device can infer the space it is currently in. In this system, RF has a role of time synchronization.

There exist systems that provide microsecond-order time synchronization on RF [32]. However, they depend on dedicated hardware or operating systems. In addition, as mentioned before, RF synchronization is not suitable for EMI-sensible situations.

2.3.3. Light Synchronization

Light synchronization for localization is an emerging technique. The concepts of the literature below are similar to our methods. They transmit light and a sound wave simultaneously, and calculate the time difference of them arrived at the receiver.

Cox et al. [5] demonstrated the proof of concept of their distance measurement method. The transmitter has an infrared LED and an ultrasonic piezoelectric sounder, and the receiver has a photodiode and a micro-electro-mechanical systems (MEMS) microphone. The transmitter begins to drive both the LED and the sounder by a 40 kHz square wave at the same time. The receiver's microcontroller calculates *TDoA* of the ultrasound and light. Then, multiplying the speed of sound, it measures the distance between the transmitter and the receiver. Receiver's signal detection is threshold-based. The experimental results showed the error of 21.6% at the 95th percentile in a dark room. They say that infrared light and ultrasonic was chosen because they detectable by human. However, they cannot be acceptable by off-the-shelf smartphones without special devices.

Rodriguez et al. [48] presented a ranging system using VLC and ultrasound. They insisted that they replaced RF in Cricket with VLC. The VLC in this system means on-off switching of a white LED. The transmitter consists of an LED driver, an ultrasound emitter, and an LED receiver. The receiver consists of an LED receiver and an ultrasound receiver. The transmitter sends a light pulse and an ultrasound signal at the same time. The receiver receives the light pulse first, then detects the ultrasound signal. When detecting the ultrasound signal, the receiver sends back a light pulse to the sender. Because the light propagation time is assumed zero in both way, both the transmitter and the receiver get the one-way ToF of the ultrasound signal. As the light detection is threshold-based, the precision is not high. In addition, this system requires dedicated devices for VLC and ultrasonic localization.

2.3.4. Summary

Table 2.4 shows the comparison of some of time synchronization methods. Network-based protocols are not suitable for smartphones because the precision depends on variable processing

time of Wi-Fi. In addition, RF cannot perform high precision under generic operating systems. Light-based systems including VLC show high performance, and they are applicable for a smartphone using its video camera. Therefore, we employ these methods for our system. A disadvantage of them is power consumption; capturing video images put a load on CPU.

Table 2.4. Localization media and methods for smartphones.

Media/Method	NTP/PTP on Wi-Fi	RF	Light	VLC
Precision	Moderate	Moderate/High	Moderate/High	High
Power consumption	High	High/Moderate	Moderate	High
Additional system cost	No	Moderate/High	Moderate	Moderate
EMI-sensible situation	No	No	Yes	Yes

2.4. Visible Light Communication

Here, we mention some VLC systems for smartphones. Some researchers have utilized CMOS image sensors for VLC and localization.

Danakis et al. [7] showed that the rolling shutter effect of CMOS image sensors could positively enhance the data rate of VLC. Capturing on–off changes of light by a smartphone's camera, they acquired striped image patterns and retrieved information from them. This system requires short-duration exposure to capture VLC messages. Therefore, it is affected by environmental noise.

Rajagopal et al. [47] used modulated LED lights as landmarks to distinguish 29 light sources in a space. Each LED light sends messages that can be used to locate the position. They demonstrated that the system worked when the user was running.

Kuo et al. [26] demonstrated *Luxapose* for indoor localization. This system identifies light sources and obtains coordinates by referring to a database. They achieved around 10 cm accuracy using a measurement based on angle of arrival.

As we noted before, VLC has an advantage that it causes no EMI. Therefore, it can be used in EMI-sensitive environments such as hospitals and plants where RF is prohibited [23].

3. 2D Time-difference-of-arrival Based Acoustic Localization for Smartphones

In this chapter, we describe the principle of an acoustic localization for smartphones. We show the feasibility of this method through 2D experiments. We also reveal the limitations of TDoA localization.

3.1. Overview of the Proposed System

3.1.1. Sound Acquisition by Smartphone

Fixed beacons send acoustic waves and a smartphone catches them. Calculating ToF of each acoustic wave, the smartphone determines its position. Acoustic waves are easily handled by smartphones. We use higher frequencies above 13 kHz, which are almost inaudible for human ear. However, smartphones can detect them. In addition, off-the-shelf loudspeakers can emit them. The signal length is to be short to reject multipath signals easily.

3.1.2. Frequency Division Multiplexing

For 2D or 3D localization, we need multiple loudspeakers. To shorten the time for localization, they send acoustic waves simultaneously. Therefore, we employ frequency division multiplexing (FDM).

3.1.3. TDoA Localization

We use standalone loudspeakers as a transmitter. Because the transmitter and the receiver smartphone are independent, they have no time synchronization. Therefore, we employ the TDoA scheme.

3.2. Frequency Division Multiplexing Phase Accordance Method

3.2.1. Phase Accordance Method

We extended PAM to frequency division multiplexing PAM (FDM-PAM) to deal with multiple sound waves simultaneously. First, we describe the *basic* PAM. Then, we conduct the FDM-PAM equation as an extension of PAM.

In PAM, a special burst pulse consisting of two acoustic waves of different frequencies is transmitted to create a beat wave called the *sync pattern*, as shown in Figure 3.1. The signal of the sync pattern can be expressed as

$$\begin{aligned} s(t) &= a_1 \sin(\omega_1 t + \phi_1) + a_2 \sin(\omega_2 t + \phi_2) \\ &= a_1 \sin(2\pi f_1 t + \phi_1) + a_2 \sin(2\pi f_2 t + \phi_2), \end{aligned} \tag{3.1}$$

where a_i is the amplitude, ω_i is the angular frequency, f_i is the frequency, and ϕ_i is the initial phase ($i = 1, 2$). The beat of the sync pattern has a time point called the *epoch* at which the phase difference of the two composing waves matches a predefined relation (zero, in this case) (Figure 3.1). The epoch is used to identify the time at which the sync pattern is transmitted from the transmitter and the time of arrival at the receiver. The carrier phases ϕ_1 and ϕ_2 at the center of the window are identified using quadrature detection.

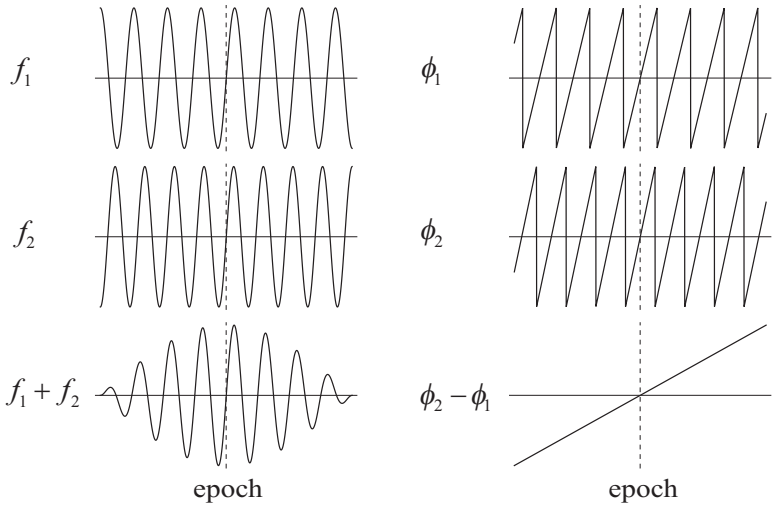


Figure 3.1. Sync pattern and its epoch.

When the sync pattern reaches the receiver, the component carriers f_1 and f_2 must be extracted and their phase information ϕ_1 and ϕ_2 must be precisely determined. To do this, we use the following mathematical process.

We define the inner product of two time-domain functions $f(t)$ and $g(t)$ by

$$\langle f(t), g(t) \rangle = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\, \overline{g(t)} dt \tag{3.2}$$

where $\overline{g(t)}$ is the complex conjugate of $g(t)$.
The inner product of $\sin(\omega t + \phi)$ and the complex exponential function

$$e^{j\Omega t} = \cos \Omega t + j \sin \Omega t, \quad \left(j = \sqrt{-1}\right) \tag{3.3}$$

yields

$$\begin{aligned}
\langle \sin(\omega t + \phi), e^{j\Omega t} \rangle &= \frac{1}{T} \int_{-T/2}^{T/2} \sin(\omega t + \phi)\, \overline{e^{j\Omega t}}\, dt \\
&= \frac{1}{2Tj} \int_{-T/2}^{T/2} \left(e^{j(\omega t+\phi)} - e^{-j(\omega t+\phi)} \right) e^{-j\Omega t}\, dt \\
&= \frac{1}{2Tj} \int_{-T/2}^{T/2} \left(e^{j(\omega-\Omega)t+j\phi} - e^{-j(\omega+\Omega)t-j\phi} \right) dt \\
&= \frac{1}{2Tj} \left(e^{j\phi} \frac{e^{j(\omega-\Omega)T/2} - e^{-j(\omega-\Omega)T/2}}{j(\omega-\Omega)} \right. \\
&\qquad \left. - e^{-j\phi} \frac{e^{j(\omega+\Omega)T/2} - e^{-j(\omega+\Omega)T/2}}{j(\omega+\Omega)} \right) \\
&= \frac{1}{2j} \left(e^{j\phi} \frac{\sin(\omega-\Omega)T/2}{(\omega-\Omega)T/2} - e^{-j\phi} \frac{\sin(\omega+\Omega)T/2}{(\omega+\Omega)T/2} \right) \\
&= \frac{1}{2j} \left(e^{j\phi} \operatorname{sinc} \frac{\omega-\Omega}{2} T - e^{-j\phi} \operatorname{sinc} \frac{\omega+\Omega}{2} T \right),
\end{aligned} \tag{3.4}$$

where

$$\operatorname{sinc} x = \frac{\sin x}{x} \tag{3.5}$$

is the sampling function. Using the above formula and observing the relations $\operatorname{sinc}(-x) = \operatorname{sinc}\ x$ and $\operatorname{sinc}\ 0 = 1$, the inner product of $s(t)$ with $e^{j\omega_1 t}$ can be described as

$$\begin{aligned}
\langle s(t),\ e^{j\omega_1 t} \rangle &= \frac{1}{2j} \left(a_1 \left(e^{j\phi_1} \operatorname{sinc} \frac{\omega_1-\omega_1}{2} T - e^{-j\phi_1} \operatorname{sinc} \frac{\omega_1+\omega_1}{2} T \right) \right. \\
&\quad \left. + a_2 \left(e^{j\phi_2} \operatorname{sinc} \frac{\omega_2-\omega_1}{2} T - e^{-j\phi_2} \operatorname{sinc} \frac{\omega_2+\omega_1}{2} T \right) \right) \\
&= \frac{1}{2j} \left(a_1 \left(e^{j\phi_1} - e^{-j\phi_1} \operatorname{sinc} \omega_1 T \right) \right. \\
&\quad \left. + a_2 \left(e^{j\phi_2} \operatorname{sinc} \frac{\omega_1-\omega_2}{2} T - e^{-j\phi_2} \operatorname{sinc} \frac{\omega_1+\omega_2}{2} T \right) \right).
\end{aligned} \tag{3.6}$$

Similarly, the inner product of $s(t)$ and $e^{j\omega_2 t}$ is

$$\begin{aligned}
\langle s(t),\ e^{j\omega_2 t} \rangle &= \frac{1}{2j} \left(a_1 \left(e^{j\phi_1} \operatorname{sinc} \frac{\omega_1-\omega_2}{2} T - e^{-j\phi_1} \operatorname{sinc} \frac{\omega_1+\omega_2}{2} T \right) \right. \\
&\quad \left. + a_2 \left(e^{j\phi_2} - e^{-j\phi_2} \operatorname{sinc} \omega_2 T \right) \right).
\end{aligned} \tag{3.7}$$

Equations (3.6) and (3.7) can be combined into a matrix equation where $a_1 e^{j\phi_1}$ and $a_2 e^{j\phi_2}$ are

the unknowns:

$$
\begin{aligned}
& i\begin{pmatrix} 1 & \operatorname{sinc}\dfrac{\omega_1-\omega_2}{2}T \\ \operatorname{sinc}\dfrac{\omega_1-\omega_2}{2}T & 1 \end{pmatrix}\begin{pmatrix} a_1 e^{j\phi_1} \\ a_2 e^{j\phi_2} \end{pmatrix} \\
& -\begin{pmatrix} \operatorname{sinc}\omega_1 T & \operatorname{sinc}\dfrac{\omega_1+\omega_2}{2}T \\ \operatorname{sinc}\dfrac{\omega_1+\omega_2}{2}T & \operatorname{sinc}\omega_2 T \end{pmatrix}\begin{pmatrix} a_1 e^{-j\phi_1} \\ a_2 e^{-j\phi_2} \end{pmatrix} \\
& = 2j\begin{pmatrix} \langle s(t), e^{j\omega_2 t}\rangle \\ \langle s(t), e^{j\omega_2 t}\rangle \end{pmatrix}.
\end{aligned} \tag{3.8}
$$

This can be strictly solved by examining the real part and the imaginary part individually.

The integration period T of the above formulas can be regarded as the duration of a square window applied to the received signal $s(t)$. Naturally, T must be chosen shorter than the length of a sync pattern.

When f_1 and f_2 are chosen as integer multiples of $1/T$, then $\omega_1 T$, $\omega_2 T$ and $(\omega_1 \pm \omega_2)T/2$ are integer multiples of π. Therefore, the terms with $\operatorname{sinc}\omega_1 T$, $\operatorname{sinc}\omega_2 T$, and $sinc(\omega_1 \pm \omega_2)T/2$ become equal to zero. Equation (3.8) then forms a direct solution as

$$
\begin{aligned}
a_1 e^{j\phi_1} &= 2j\langle s(t), e^{j\omega_1 t}\rangle \\
a_2 e^{j\phi_2} &= 2j\langle s(t), e^{j\omega_2 t}\rangle.
\end{aligned} \tag{3.9}
$$

We find such a choice of T in the receiving process of orthogonal frequency division multiplexing (OFDM) communication.

In any case, phase and amplitude information of the f_1 and f_2 components can be obtained from the sync pattern by the multiplication and integration of the received signal with $\overline{e^{j\omega_1 t}}$ and $\overline{e^{j\omega_2 t}}$. The multiplication of the received signal of carrier angular frequency ω with $\cos\omega t$ and $\sin\omega t$ is called *quadrature detection*, and is widely used for the demodulation of phase-modulated signals. Through the above equation, the same method can be applied to obtain precise phase information from multiple carrier signals. The integration interval T is called a *window*. Note that when the received signal $f(t)$ is sampled and converted to a digital number series f_n the inner product of $f(t)$ and $g(t)$ is

$$
\langle f(t), g(t)\rangle \approx \frac{1}{N}\sum_{i=1}^{N} f_i\,\overline{g_i}. \tag{3.10}
$$

With this algorithm, we can obtain the carrier phases ϕ_1 and ϕ_2 at the center of the window T, and set it as $t = 0$. Assuming $\omega_1 < \omega_2$, ϕ_2 increases faster than ϕ_1. As time passes, ϕ_2 catches up with ϕ_1 and eventually the two phases become identical again. This happens at

$$
t = -\frac{[\phi_1 - \phi_2]}{\omega_1 - \omega_2} = -\frac{[\phi_1 - \phi_2]}{2\pi(f_1 - f_2)}. \tag{3.11}
$$

$[x]$ denotes that the phase value is described with its principal value $-\pi < x \leq \pi$. The name *phase accordance method* comes from this algorithm for detecting the epoch when the two carriers form a certain difference (identical values in this case). The delay t can be converted to the distance using the known propagation speed c, which is a ToF measurement.

Advantages of PAM

For PAM, we usually choose a pair of f_1 and f_2 with close frequencies. In ultrasonic measurements, for instance, $f_1 = 39.75\,\text{kHz}$ and $f_2 = 40.25\,\text{kHz}$, then $\Delta f = f_2 - f_1 = 500\,\text{Hz}$. This means that the transducers (loudspeakers and microphones) for PAM only deal with a narrow bandwidth, therefore reducing the requirements for the transducers. PAM uses only phases of the signal. Phase characteristics are preserved better than amplitude in signal transmission. Thus, very accurate distance measurements, of sub-millimeter standard deviations, can be achieved.

Other types of phase interferometry such as Cricket Compass [46] also work in the same manner. However, simple phase interferometry gives multiple candidate measurement values separated by a wavelength, and some additional selection algorithms are required to determine the true value. PAM does not require the selection process as it gives a single distance value.

Because of the high accuracy of PAM distance measurements, only short baselines (e.g., a few centimeters) are needed for multilateration. We can implement equipment for PAM multilateration by mounting multiple loudspeakers (or microphones) on a compact box. For localization, we simply put the box on a shelf instead of deploying individual transducers on the room walls; this greatly simplifies the deployment process.

A disadvantage of PAM is that it is sensitive to Doppler effects; therefore, measuring a fast-moving object is often difficult. To overcome this problem a new algorithm *Extended Phase Accordance Method* (EPAM) [43, 52] has been developed. The velocity of the object can be detected with high accuracy using EPAM.

Effect of Multipath

For indoor localization systems using modulated waves (e.g., measurements using pseudorandom noise sequences) the troublesome phenomenon is multipath. The reflected waves are received with the main (direct) signal and interfere with the measurements.

Figure 3.2 shows an example of the multipath characteristics observed in an office room. Such characteristics are also called the reverberation of a room. After the arrival of the direct-path signal, many multipath signals (reflections or echoes) follow it. The reverberation is sustained for almost 100 ms. The measured high-frequency range 14–17 kHz usually has shorter reverberation; however, the effect is still quite strong.

One way to avoid multipath interference is to perform a very fast measurement using short-duration signals. When we look closely at the delay profile (Figure 3.2), we find a gap of about 8 ms between the direct path and the first multipath. In this case, the first multipath seems to be caused by the reflection on the ceiling. If we use a shorter wave packet than the gap time, we can perform localization without suffering multipath interference.

This was our first motivation for trying the phase accordance method (PAM) for the indoor positioning of smartphones. In PAM, which we developed and applied for ultrasound localization, the detection wave is only a few milliseconds long.

FDM-PAM

To locate a mobile receiver based on multilateration, we send out multiple sync patterns from fixed beacons. As the signal sources must be discriminated on the receiver side, multiplexing technology (e.g., time domain or frequency domain) must be employed.

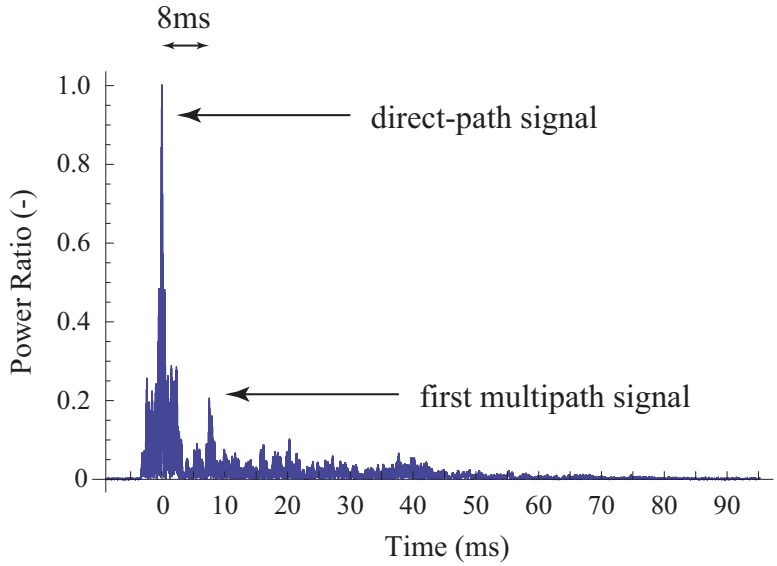


Figure 3.2. Reverberation—delay profile of 14–17 kHz sound wave.

If we apply *time division multiplexing* (TDM) for indoor localization, we must consider multipath of the room as mentioned before. As have seen in Figure 3.2, it takes around 100 ms for the diminution of the reverberation and only after that is the successful transmission of the PAM measurement possible. This makes localization using TDM slow.

In contrast, when we adopt FDM-PAM, simultaneous transmissions of sync patterns from different beacons can accomplish the localization process. If the sync pattern sits in the gap between the direct path and the first arrival of a multipath signal, the localization can be achieved without suffering from interference.

A smartphone localization algorithm that uses only the ToF of sound is based on the TDoA scheme. For TDoA localization the number of beacons required is the number of dimensions plus one, i.e., 3 for 2D localization and 4 for 3D. This means six carrier frequency channels are required for a 2D FDM-PAM system. Let us consider such an extension of the PAM algorithm using six frequencies.

Let $f_1, f_2, \ldots, f_6$ be the six frequencies separated with a constant frequency separation Δf. Three sync patterns will be formed coupling the adjoining frequencies. The matrix equation (3.8) can be extended to an arbitrary number of carriers including six. With a shorter window size T, the condition number of the matrix increases rapidly as the number of carriers increases. For multiple-frequency PAM with more than two carriers we require the window size $T \geq 1/\Delta f$. We must therefore set $T = 1/\Delta f$, which also provides the simplest detection formula (3.9). However, for this procedure we must identify the overlap part of the three sync patterns that is longer than T, so naturally all the sync patterns must be longer than T.

Here we use $2T$-long sync patterns for FDM-PAM. They have two beats as shown in Figure 3.3. We place the epoch at the center of the beats (at the nodes) instead of the antinodes. As the two frequency carriers are combined destructively at the nodes, the epoch (node) is detected rather differently

$$t_1 = -\frac{[\phi_1 - \phi_2 - \pi]}{\omega_1 - \omega_2}, \quad etc. \tag{3.12}$$

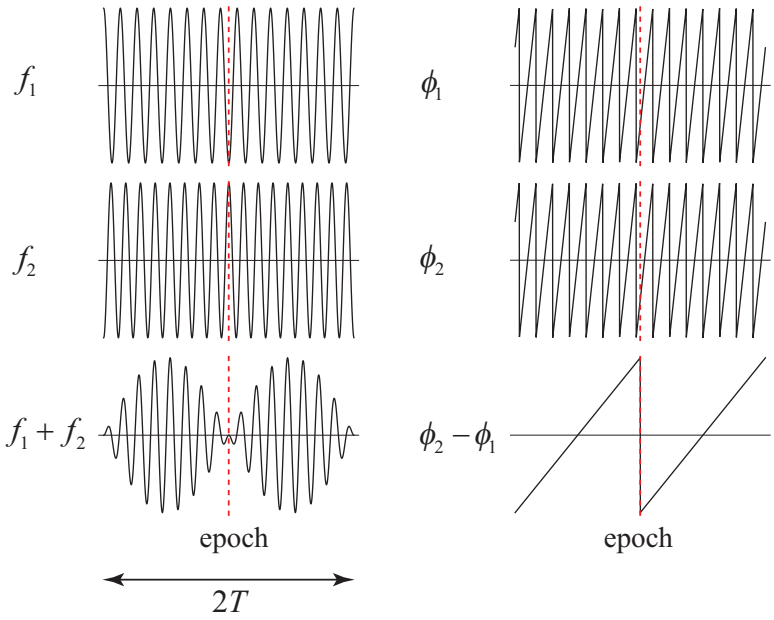


Figure 3.3. $2T$ sync pattern and its epoch.

instead of the approach in basic PAM, which is (3.11). There are still two other nodes on the $2T$ sync pattern on the both sides, but they are easily distinguished and eliminated when we apply the window to the signal.

The overlap of three sync patterns longer than T is only available in a limited zone in a room when the beacons transmit them simultaneously. Assuming the three beacons are mounted in a horizontal line separated 0.5 m from each other, the area suitable for the FDM-PAM is shown in Figure 3.4. We assume that for most applications, we can arrange the positions of beacons as well as their separation to cover the required service area.

3.3. Precision Evaluation by Noise Simulation

In basic (two-frequency) PAM it is known that the variance of a ToF measurement ${\sigma_t}^2$ is theoretically described as

$$(2\pi\Delta f\sigma_t)^2 = \frac{2}{1 - \mathrm{sinc}(\omega_1 - \omega_2)T/2}\frac{W}{PT}, \tag{3.13}$$

where P is the power of a sync pattern and W is the power of the noise (white noise is assumed) [16]. More particularly, the standard deviation ${\sigma_t}^2$ is defined as proportional to W/PT, which is the inverse of the *SNR* value.

For multiple-frequency FDM-PAM we have not yet analyzed the noise characteristics theoretically; however, a similar result is expected. To confirm it we conducted a noise simulation. The carrier frequencies $f_1, f_2, \ldots, f_6$ were chosen as 14.75, 15.25, . . ., 17.25 kHz and three sync patterns s_1, s_2 and s_3 were formed using adjoining carriers. The period of a sync pattern is 4 ms and window size $T = 2$ ms. This window size is short enough to avoid multipath interference, and long enough to collect all the sync patterns. The amplitudes of the three sync patterns were

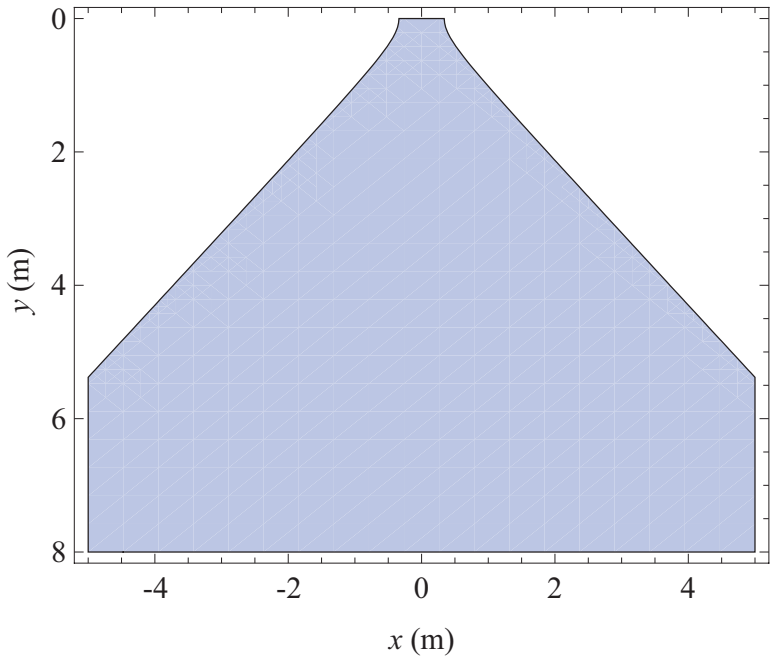


Figure 3.4. Zone coverage of $2T$ sync pattern overlap $> T$.

set at 1 V, 0.5 V and 0.33 V respectively. We also provided 0.2 ms time-offsets between the sync patterns.

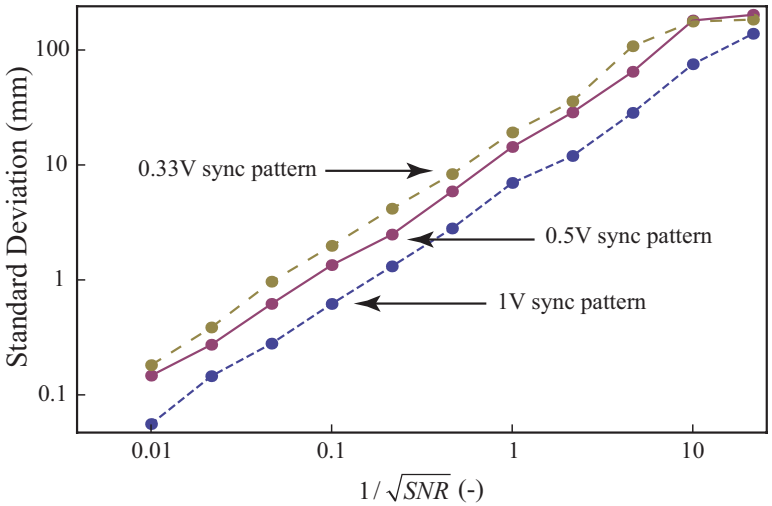


Figure 3.5. Expected error with white noise.

Figure 3.5 shows the simulation result. The horizontal axis describes the $1/\sqrt{SNR}$ value. We set the signal sampling rate at 48 ksps. White-noise power is assumed spread evenly in the range of 0 kHz to 24 kHz. The vertical axis shows the expected standard deviation of the measurement using the sound velocity of 340 m/s. As we expected, the simulation graph shows that the precision of FDM-PAM depends on the factor $1/\sqrt{SNR}$. Another interesting point in this simulation is that the error value is distributed to each sync pattern depending on the inverse of the component carrier amplitudes.

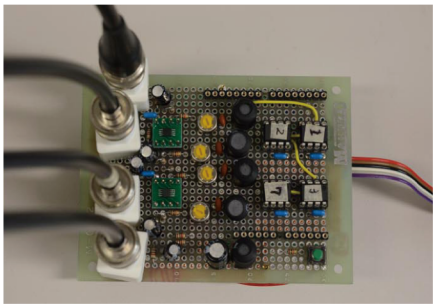
3.4. 2D TDoA Localization Experiment using FDM-PAM

3.4.1. Experimental Setup

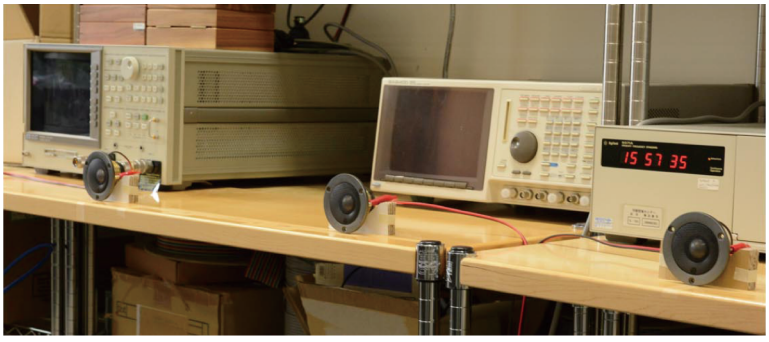
We conducted an experiment to confirm the feasibility of FDM-PAM in real indoor situations. For the sync-pattern transmitters we used three tweeters (FT200D, FOSTEX) driven by a lab-made multichannel waveform generator, shown in Figure 3.6. For the receiver we used a smartphone (XPERIA acro HD SO-03D, SONY Mobile Communications), which recorded the signal from the tweeters located 2.1 m distant from the closest speaker. In this experiment we performed the FDM-PAM signal processing using an external PC instead of using the smartphone, however, it will be implemented as a phone program later to realize stand-alone, real-time localization applications.

The system was laid out in a 3.5 m × 7.3 m × 2.8 m (height) room shown in Figure 3.7. All components were placed at the same level (0.70 m) above the floor, which allows a 2D localization. By setting the peak driving power of the speakers to 0.016 W they generated circa 68 dB sound pressure at 1 m distance. Because of its high frequency and comparatively low transmission power, the audio signals from the speakers were almost inaudible to people in the room despite the quiet environment.

Receiver Phone SO-03D,
Sony Mobile Communications



4ch Waveform Generator, home-made



Transmitter Speakers (Tweeters) 3 × FT200D, Fostex

Figure 3.6. Experimental system for 2D TDoA localization.

In the arrangement of Figure 3.7, the distance from the phone to the closest beacon was 2.10 m, and the distance to the farthest one was 2.33 m. The time difference of their sonic propagation time was thus 0.665 ms. As this is shorter than the half length of the sync pattern (2 ms), all three sync patterns should arrive at the receiver with more than T =2 ms overlap.

The signal processing algorithm in the receiver detected the presence of the patterns as well

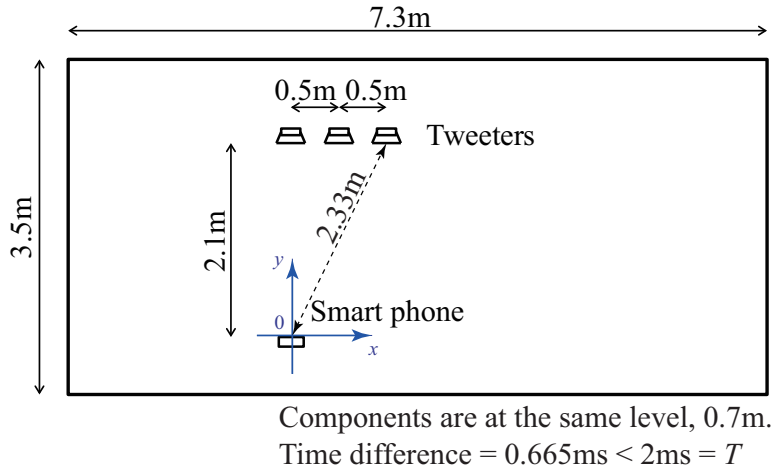


Figure 3.7. Layout of speakers and phone in experiment room.

as the overlap length using a correlation operation. The received signal in Figure 3.8 shows the positions at which the three sync patterns from the different sources were superimposed on each other. The *SNR* of this signal was calculated as 323.4 = 25.1 dB.

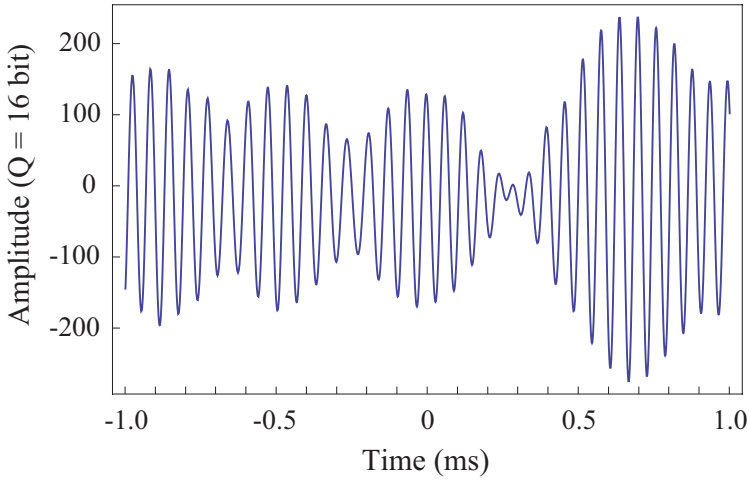


Figure 3.8. Received signal extracted with $2T$ window.

We tried FDM-PAM using the same frequencies as the simulation. The phone was first placed at the origin point (0 mm, 0 mm), then was moved, in parallel with the speakers, to the point (200 mm, 0 mm). At each point $N = 100$ localization operations were performed for statistical analysis.

3.4.2. Results

The results are shown in 3.1. As the mean value, the phone at the origin point (0, 0) was detected at (21.16, 102.81), and at (200, 0) was detected as (211.03, 13.26). The offset error was thus around 10 cm range.

Table 3.1. 2D TDoA localization results.

Position (mm)	Mean (mm)	Standard deviation (mm)		Offset
(x, y)	(x, y)	(x, y)	Overall	(mm)
(0, 0)	(21.16, 102.81)	(24.96, 110.7)	113.5	105.0
(200, 0)	(211.03, 13.26)	(26.18, 150.8)	153.1	17.2

Figure 3.9 shows the distribution of the localization results. The dots enclosed by the left ellipse show the distribution when the phone was placed at the "+" symbol at coordinates (0, 0). Similarly, the dots enclosed by the right ellipse show the distribution of the phone at (200, 0). We notice the noncircular, prolonged shape of the distributions. This is because the FDM-PAM measurement is based on TDoA, hence the locations are calculated as the crossing point of hyperbolic lines having the foci at the sound sources. A unit error of ToF measurement gives much greater effect on the radial (range) direction measurement than on the lateral direction.

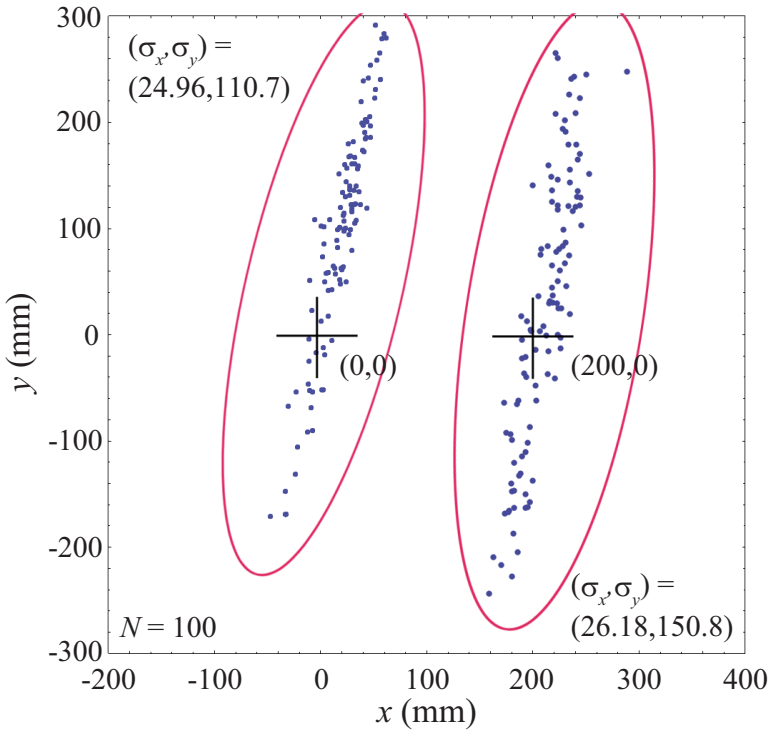


Figure 3.9. Localization results of phone at (0, 0) and (200, 0).

A comparison between TDoA and ToA is illustrated in Figure 3.10. In the TDoA chart, there are four hyperbolic lines that indicate the same TDoA. Two pairs of the four lines have

the same foci, and two lines in a pair express error bounds (±10 ms). The area surrounded by the four hyperbolae is the estimated region containing the phone. Its shape corresponds to the distributions in Figure 3.9. On the other hand, in ToA, we use just two speakers, and the four lines are circles that indicate the same ToA. The surrounded area is smaller and not prolonged. This shows that ToA's dilution of precision (DOP) is better than that of TDoA. If we enlarge the interval of speakers, we can improve the DOP in TDoA. However, the measurable area in a room decreases and deployment becomes more difficult.

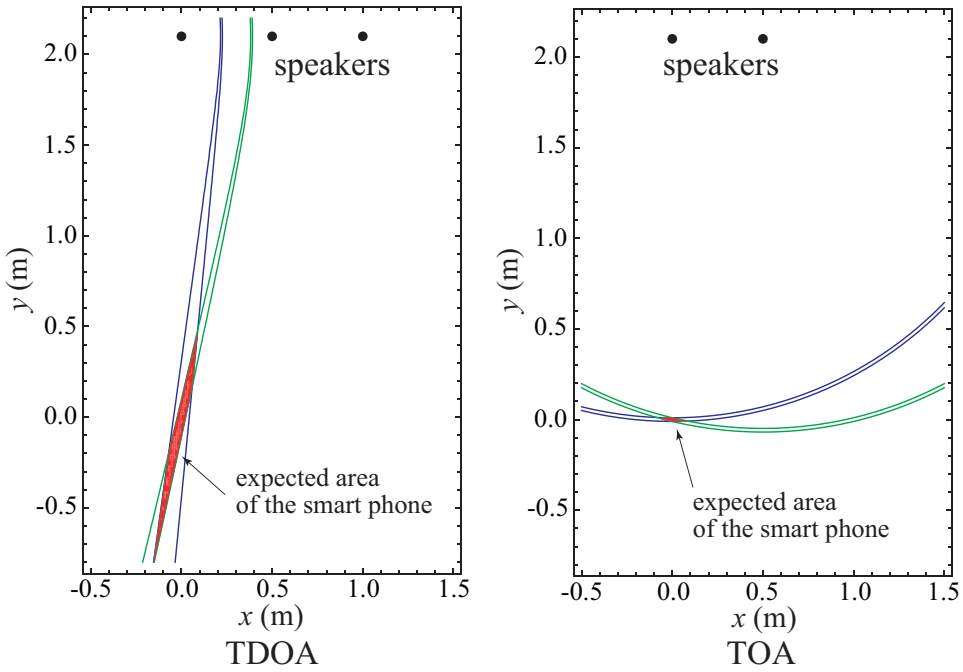


Figure 3.10. Comparison between TDoA and ToA.

Reflecting the uneven distribution of the localization error, the standard deviations of the measurements were (24.96 mm, 110.7 mm) for the object at (0, 0), and (26.18 , 150.0 mm) for (200, 0). When we estimate the ToF measurement error between a speaker and the receiver using the *SNR* value it will be less than 1 mm. The localization error mentioned above, however, should come from ToF errors of around 3 mm.

3.4.3. Effect of Environmental Noise

A possible additional error source is the environmental (ambient) noise at the experiment site. To evaluate the effect of noise we collected railway station noise in a crowded station hall, then at the laboratory we mixed it with the FDM-PAM measurement signal and checked the effect. The power ratio of the signal and the noise was set at 1 : 1 to make a 0 dB signal sample.

The result is shown in Tabletab:2d-loc-env-noise. The measurement errors were somewhat increased; however, the increment of the error is not very significant considering the low *SNR* value was used (i.e., simulating a noisy environment).

From this experiment, we concluded that the environmental noise was not the major error source. Moreover, FDM-PAM is quite tolerant of natural environmental noise. The detection process of PAM utilizes quadrature detection, which has very sharp band-pass filter (BPF)

Table 3.2. Simulation of the environmental noise at (0, 0).

	Mean (mm)	Standard deviation (mm)		Offset
	(x, y)	(x, y)	Overall	(mm)
Experiment	(21.16, 102.8)	(24.96, 110.7)	113.5	105.0
With environmental noise	(20.22, 117.4)	(28.45, 124.8)	128.0	119.1

functionality. When the environmental noise does not contain the highest audible frequency range, then FDM-PAM is virtually immune to it. This is the case for the noise of our daily life; however, it will not hold for some kinds of artificial noise that contain high-frequency sounds.

3.5. Summary

We described an acoustic localization system using smartphone. To measure precise ToF of multiple acoustic waves simultaneously, we developed FDM-PAM, an extension of PAM. A localization process can be carried out by receiving a sync pattern lasting only 4 ms in our case, enabling fast localization. Multipath interference can also be avoided because of the short duration of the sync pattern.

With simulations and an experiment, we confirmed the operation of FDM-PAM and its applicability to indoor localization. However, the experimental results contained larger errors than were expected from the simulation. We should pursue the unknown error source to improve the performance of FDM-PAM.

4. 3D Time-of-Arrival Based Acoustic Localization Using Light Synchronization

4.1. ToA and Time Synchronization

In Chapter 3, we introduced 2D localization of smartphones using three fixed acoustic beacons and a target phone using TDoA (Figure 4.1). The accuracy obtained was around 10 mm in the azimuth direction and 100 mm in the range direction. Because this uneven distribution of errors was caused by the characteristics of TDoA, it can be improved if we switch to ToA measurements to synchronize a system.

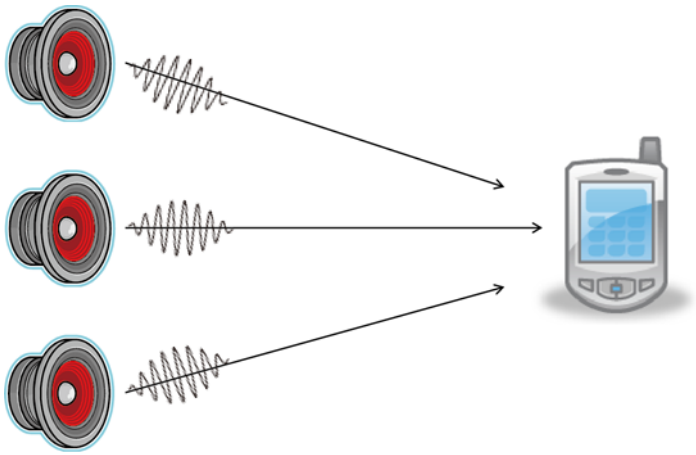


Figure 4.1. TDoA acoustic localization in 2D arrangement.

In this chapter, we describe ToA localization for smartphones. We use LED light for time synchronization for ToA localization (Figure 4.2). Light travels much faster than sound. Assume that the transmitter emits light and sound waves at the same time. If the smartphone detects the time when the light turns on by its clock, we can consider it as the emission time of the light and sound waves. This is the time synchronized situation in this study.

We use modulated LED light for the synchronization. A CMOS image sensor reads out pixel values line by line. Therefore, obtained image is modulated when the lighting of the field is dynamically modulated. When the LED light is modulated, the image from the sensor is cross-modulated by the light modulation and the sensor's scanning modulation. By analyzing the image, it is possible to synchronize the system and the node by observing some picture frames. We describe a new method for synchronizing the beacon system and the target smartphone using the properties of the CMOS image sensors that are now widely available in mobile devices.

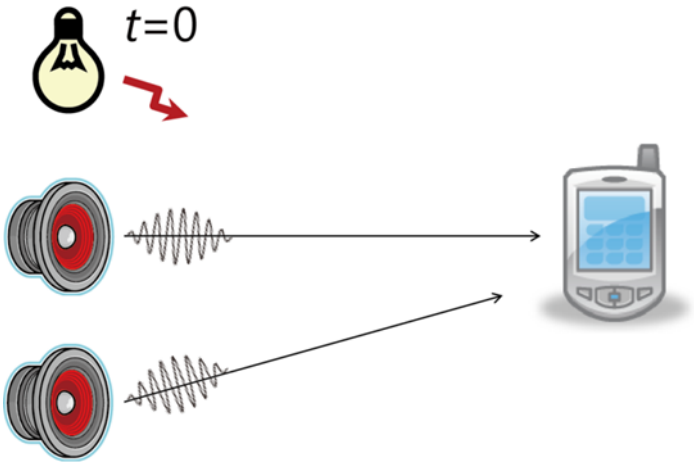


Figure 4.2. ToA acoustic localization using light synchronization in 2D arrangement.

4.2. Proposed System

4.2.1. Rolling Shutter Effect

CMOS image sensors are widely available in recent video cameras and smart devices. An image sensor has a 2D array of photodiodes that records a meshed image as electrical charges. The sensor periodically reads out the charges, and reconstructs the original image as an array of pixels. Because CMOS image sensors scan the horizontal lines of the array one by one, each line in an acquired image records the scene at slightly different times. Therefore, when we take a picture of a moving object with a CMOS image sensor, we obtain a distorted image of the object if it moves rapidly enough (Figure 4.3). This is called the *rolling shutter* effect.

The rolling shutter effect is usually considered as undesirable. However, we can make use of it as a useful measure to record events occurring in the relevant time range [1, 28].

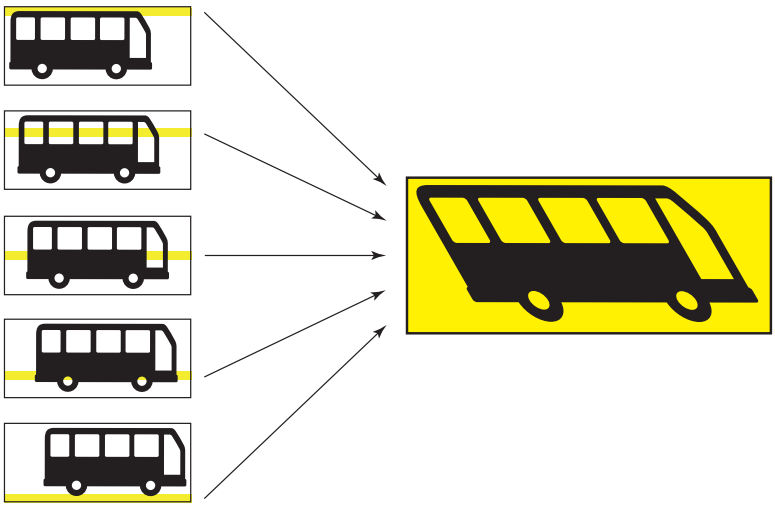


Figure 4.3. Rolling shutter effect.

4.2.2. Capturing Light with Video Cameras

We modulate an LED light by a square wave and capture it with a CMOS video camera. We suppose that the modulation frequency is exactly half of the frame rate of the video camera. Consequently, the frame interval (the reciprocal of the frame rate) is the same as the duration for which the LED is on or off. In this condition, we can detect gradation patterns of the lighting in the acquired images.

Figure 4.4 illustrates the mechanism of the gradations. We assume that the line's exposure time is the same as the frame interval, and the CMOS image sensor scans row lines across the frame one by one. If a line's exposure is coincident with the total period of the LED's activation, that line is the brightest in the frame.

The left part of Figure 4.4 (b) shows the profile of the pixel values in the left frame of Figure 4.4 (a). If the LED light illuminates the sensor cells uniformly, all the cells of the line have the same output. Therefore, we can use the average output of the line instead of treating the individual cells of the line. The horizontal axis denotes the row position in the frame. The vertical axis expresses the ratio of the time when the LED light is on in that line to the exposure time. We call this frame an *A-shaped frame* as it has a peak in it. The peak corresponds to the point of time when the LED light turns on.

Similarly, if a line's exposure is coincident with the total period when the LED light is off, that line is the darkest in the frame as shown in the right part of Figure 4.4 (a). We call this frame a *V-shaped frame*. The deepest point in the graph shows the point at which the LED light turns off. A-shaped frames and V-shaped frames appear in an alternating sequence.

There is dead time between the last row of one scan and the first row of the next frame. If the LED light turns on or off during the dead time, neither a peak nor a trough exists in a frame and the profile monotonically increases or decreases. The dead time makes the profile $p(x)$ in Figure 4.4(b) discontinuous. If we assume virtual rows that are equivalent to the dead time between frames, we should get a triangle wave with no leap at the frame boundaries. We have also demonstrated that the luminance distribution when the image sensor captures light can be modulated by an arbitrary function (see Appendix A for details).

4.2.3. Extracting Time Reference from Video Images

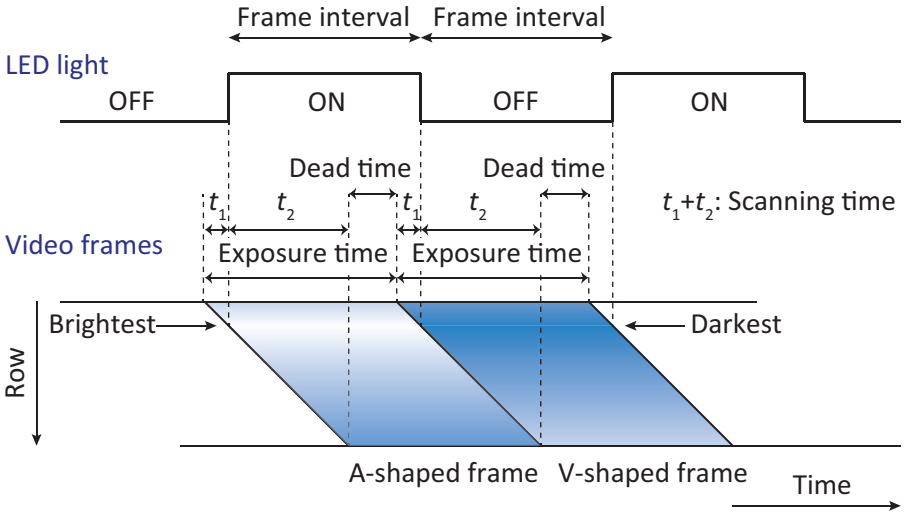
We obtain the positive peaks of the A-shaped frames or the negative peaks of the V-shaped frames as a time reference. Figure 4.4(c) shows the luminance distribution $f(x)$ over the sensor lines when the LED light is always on. This is theoretically constant, but must be adjusted by functions describing uneven LED light distributions, characteristics of the lenses, and variations of sensor cell sensitivity. Given that the profile of the A-shaped frame is $p(x)$, the distribution profile of the V-shaped frame should be $1 - p(x)$, where x is the row position in the frame.

Figure 4.4(d) shows frame pixel values, $A(x)$ in the A-shaped frame and $B(x)$ in the V-shaped frame. Dashed lines are the background luminance distribution $b(x)$ when the LED light is off.

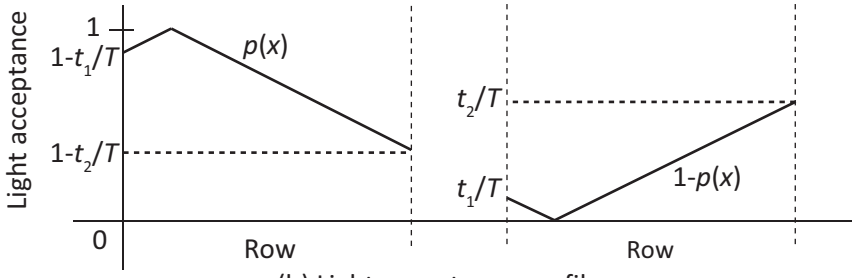
We can describe $A(x)$ and $B(x)$ as follows:

$$\begin{aligned} A(x) &= p(x)f(x) + b(x), \\ B(x) &= (1 - p(x))\, f(x) + b(x). \end{aligned} \tag{4.1}$$

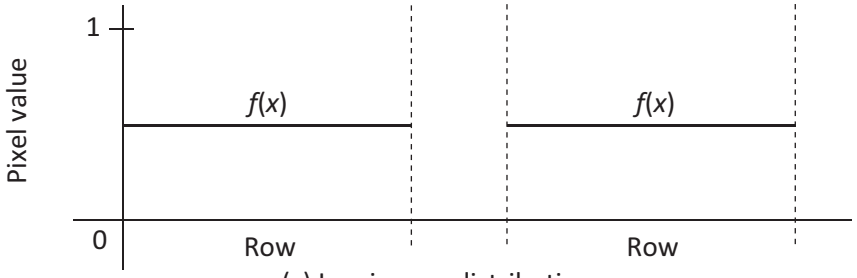
Here, $A(x)$ and $B(x)$ are measurable values; $p(x)$, $f(x)$, and $b(x)$ are unknowns. Assuming $f(x)$ or $b(x)$ for some value (e.g., setting $b(x)$ to the minimum pixel value of $B(x)$), we can calculate



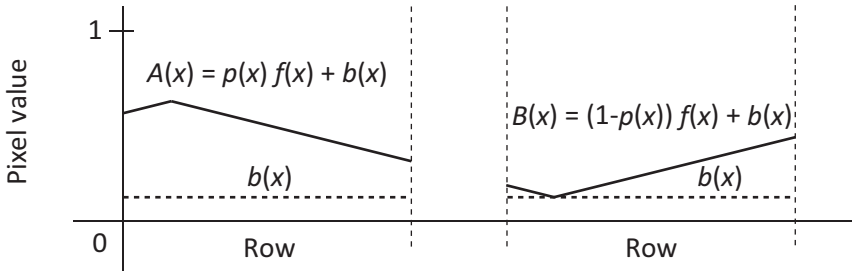
(a) Mechanism for occurrence of gradations



(b) Light acceptance profile



(c) Luminance distribution



(d) Frame pixel values and background luminance distribution

Figure 4.4. Mechanism for occurrence of gradations.

the remaining unknowns. Then we obtain the time reference as the maximum position of $p(x)$.

The maximum position of $p(x)$, which can be obtained from the maximum position of $A(x)$, is calculated as follows: $A(x)$ consists of two straight lines that intersect at and are symmetrical about the maximum position of $A(x)$. Pixel values of rows obtained from the captured frame are divided into two groups separated by the maximum position. Then, they are adapted to two regression lines that form an A-shape; the intersection of the regression lines is the precise maximum position, which has a resolution finer than the row interval.

4.2.4. Solving 3D ToA and TDoA

In the 3D arrangement, the estimated smartphone position is one of the intersections of three or more spheres or hyperboloids. However, in some cases there is no intersection because of measurement errors. This means that there is a possibility of obtaining no real number solutions around the interested area for the simultaneous equations that represent the surfaces of spheres or hyperboloids. Therefore, we describe the ToA and TDoA calculation algorithms as the following optimization problems because they have two types of solutions: the simultaneous equation solutions (there exist intersections and the minimum value becomes zero) and the approximation solutions (there are no intersections and the minimum value remains non-zero).

For ToA localization, we solved

$$\arg\min_{\boldsymbol{x}} \sum_{i=1}^{n} \|\boldsymbol{p_i} - \boldsymbol{x}\|^2 \quad (n \geq 3) \tag{4.2}$$

such that

$$\|\boldsymbol{p_i} - \boldsymbol{s_i}\| = r_i, \tag{4.3}$$

where n is the number of loudspeakers; $\boldsymbol{p_i}$ is a position vector that indicates a point on the sphere with center $\boldsymbol{s_i}$ and radius r_i; $\boldsymbol{x}$ is the smartphone's position to be determined; $\boldsymbol{s_i}$ is the position of the ith loudspeaker; and r_i is the distance between the ith loudspeaker and the smartphone that is derived from the product the measured ToF and the speed of acoustic waves. Furthermore, $\|\cdot\|$ denotes the Euclidean norm.

For TDoA, we solved

$$\arg\min_{\boldsymbol{x}} \sum_{\substack{i=1 \\ i\neq j}}^{n} \|\boldsymbol{p_{ij}} - \boldsymbol{x}\|^2 \quad (1 \leq j \leq n,\ n \geq 4) \tag{4.4}$$

such that

$$\|\boldsymbol{p_{ij}} - \boldsymbol{s_i}\| - \|\boldsymbol{p_{ij}} - \boldsymbol{s_j}\| = c(t_i - t_j), \tag{4.5}$$

where $\boldsymbol{p_{ij}}$ is a point on the hyperboloid whose TDoA to loudspeakers $\boldsymbol{s_i}$ and $\boldsymbol{s_j}$ is $c(t_i - t_j)$, c is the speed of sound, and t_i and t_j are the smartphone clock times when sound waves from the ith and jth loudspeakers arrived, respectively. Note that the jth loudspeaker is the reference against which TDoA is calculated.

4.3. Time Synchronization Experiment

We conducted an experiment to confirm the feasibility of the time synchronization method. Using the function generator's trigger signal as the ground-truth time reference, we examined the precision of the experiment. In addition, 1D ranging measurements were performed.

4.3.1. Experimental Setup

Figure 4.5 shows the diagram of the experiment system. We used the LED floodlight shown in Figure 4.6, which consists of 56 infrared LEDs (OSI5LA5113A, OptoSupply; peak wavelength: 940 nm). Although infrared light is invisible to the human eye, some video cameras are sensitive to it. The floodlight was covered by thin tracing paper to diffuse the directive rays emitted by LED elements. The light was modulated by a 29.97 Hz square wave generated by a function generator (AFG3102, Tektronix). The function generator also generated a sound wave, which was emitted by a tweeter (FT200D, FOSTEX). A video camera (HDR-CX700V, SONY) took video images of the LED floodlight. Each image has 1920×1080 pixels. The video frame rate is 59.94 fps (progressive), and the audio sampling rate is 48 ksps. An electret condenser microphone (ECM-G5M, SONY) was connected to the video camera to record the acoustic sound emitted by the tweeter. The microphones attached to the video camera were not used. The electric sound wave signal generated by the function generator was also recorded by the video camera. This signal was used to validate the time reference found by the frame analysis and to perform acoustic localization for evaluating the precision.

The video camera's uncompressed HDMI output was transmitted in real time to a video capture board (MonsterX U3.0R, SKnet), and the signal was recorded in an uncompressed multimedia format. After that, using an FFmpeg tool (http://www.ffmpeg.org/), the contents were split into image files and an audio file to be analyzed separately.

Figure 4.7 shows the square wave modulating the LED floodlight and the audio signal. The acoustic pulses generated for the ToA measurement are synchronized to the rising edge of the square wave driving the LEDs once in every six cycles. This acoustic pulse is a sync pattern, which consists of two sinusoidal waves of frequencies 14.75 kHz and 15.25 kHz. Using FDM-PAM, we can precisely determine the epoch, the center of the sync pattern.

The experiment was performed in a darkened office room. The tweeter and the microphone were placed at the same level (68 cm) above the floor, and their separations were chosen as 1.0 m, 1.5 m, 2.0 m, 3.0 m, and 4.0 m. By setting the peak driving power of the tweeter to 0.016 W, it generated circa 68 dB sound pressure at 1 m distance. We took measurements five times for each distance.

The video camera and the LED floodlight were also placed at the same level (77 cm) above the floor, and their separation was 60 cm. The peak driving power of the LEDs was 7.6 W, and the radiant power is estimated to be 1.3 W. The video camera's capture time was five seconds or more for each measurement.

4.3.2. Results

The acquired data were analyzed as follows. We coupled adjoining image files, which consist of an A-shaped frame and a V-shaped frame. By averaging the pixel values in each row, we created the vectors $A(x)$ and $B(x)$ for the A-shaped frame and the V-shaped frame, respectively.

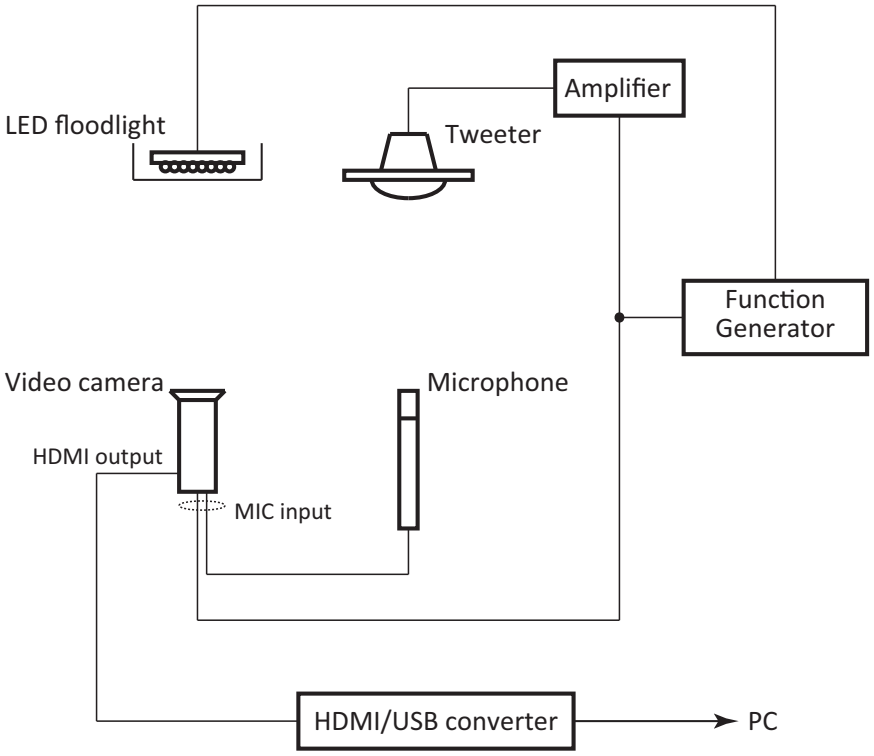


Figure 4.5. Experimental system for time synchronization.

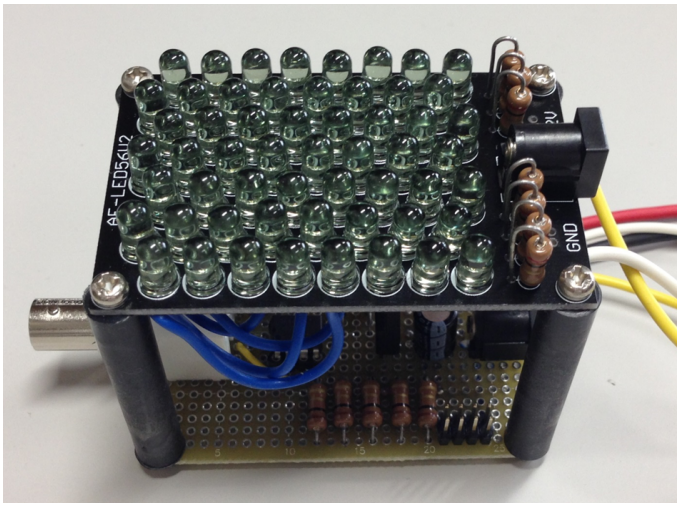


Figure 4.6. Infrared LED floodlight.

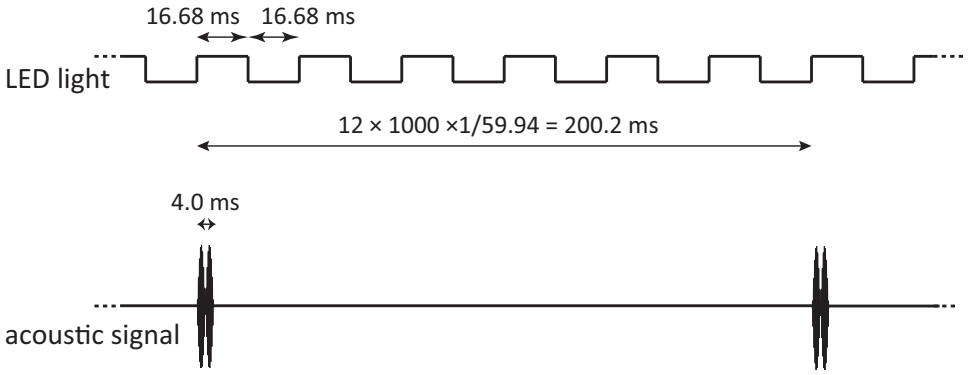


Figure 4.7. LED modulation pattern and audio signal.

When the background luminance distribution $b(x)$ was set at the minimum pixel value of $B(x)$, we could solve (4.1) to determine $p(x)$ and $f(x)$. Consequently, we obtained the time reference point x at which $p(x)$ has the maximum value.

Figure 4.8 shows the raw data taken during the experiment. Figure 4.8 (a) shows the A-shaped vector $A(x)$ and the V-shaped vector $B(x)$, which are actually curved. Figure 4.8 (b) describes the profile $p(x)$ and the LED's luminance distribution $f(x)$. Fitting $p(x)$ to two straight lines, we obtained the time reference point. $f(x)$ shows a nonuniform distribution.

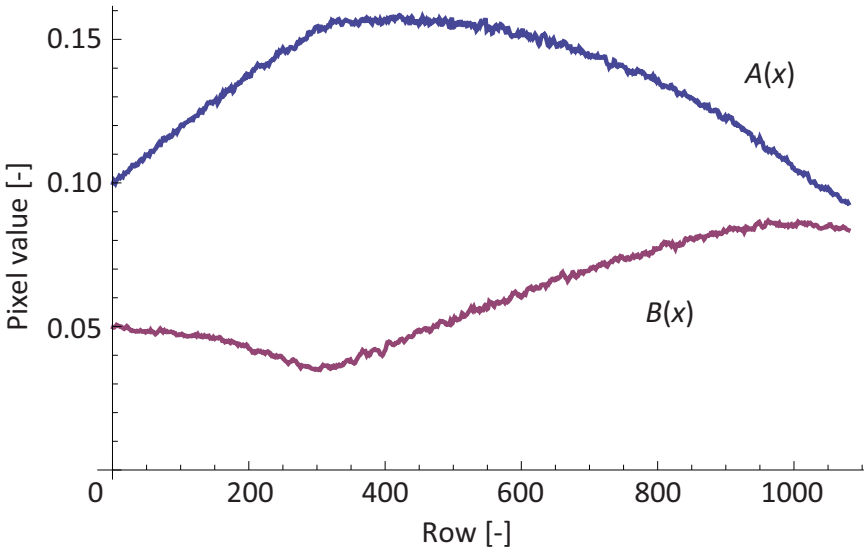
In 15 cases of distances of 1.5 m, 3.0 m, and 4.0 m, we could obtain the time references. The time reference points calculated by the frame pairs fluctuated. However, by averaging at most 27 frame pairs (54 frames, i.e., within one second), the time reference point had a standard error of less than 0.5 rows.

The audio file contains the electric signal produced by the function generator and the acoustic signal through the microphone. We recorded the raw electronic signal source for the loud speaker as well as the sound captured by the microphone. Using FDM-PAM, the first wave's position in the electric signal was determined. As this can be accomplished in a purely electronic process, the timing of the position was obtained with very high accuracy, in the 0.5 μs range. We used this as the time reference extracted from the audio file. Then we regarded the electric signal as the time origin and calculated the ToF of the acoustic signal, and achieved a 1D localization process. The standard deviations observed in the acoustic ToF were less than 2 mm for the 15 cases of distances 1.5 m, 3.0 m, and 4.0 m.

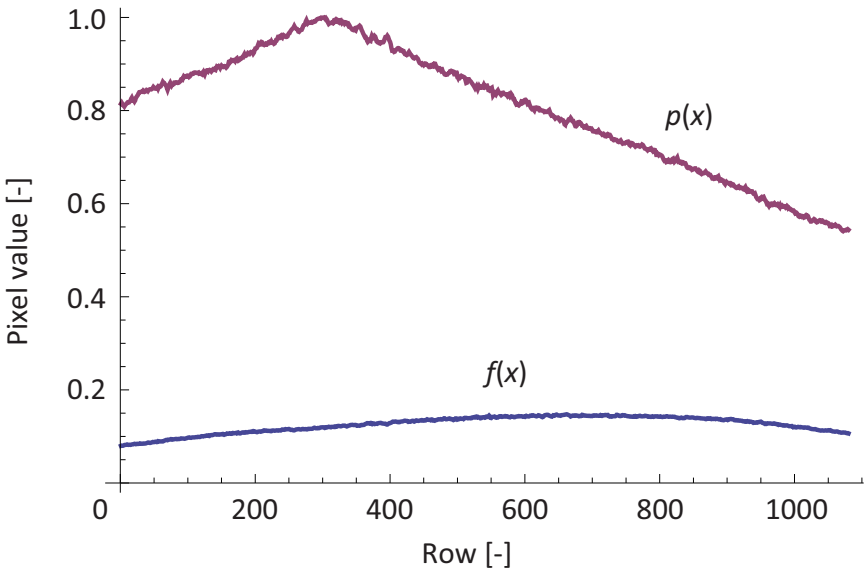
By comparing the time references from the image files and the audio file, we confirmed the accuracy of our method. Figure 4.9 shows the relations between the electric timings and the synchronization reference obtained from the image file. All 15 data points clearly sit on a line. The regression equation is

$$y = -3.75 \times 10^{-5} + 8.84 \times 10^{-6} x, \tag{4.6}$$

where x is the time reference point at which $p(x)$ becomes a maximum, and y is the position at which the first electric signal exists in the same multimedia file. The standard deviation of this regression is 0.017 ms, which is equivalent to 1.9 lines in the image, and sound travels about 5.8 mm during this time. This value is rather worse than our previous assessment, which was around 0.5-line time based on the variance analysis. This might be caused by an alignment mismatch between the video and audio in a frame.



(a) Obtained pixel values in A-shaped frame ($A(x)$) and V-shaped frame ($B(x)$)



(b) Estimated light acceptance profile $p(x)$ and luminance distribution $f(x)$

Figure 4.8. Sample of image analysis process.

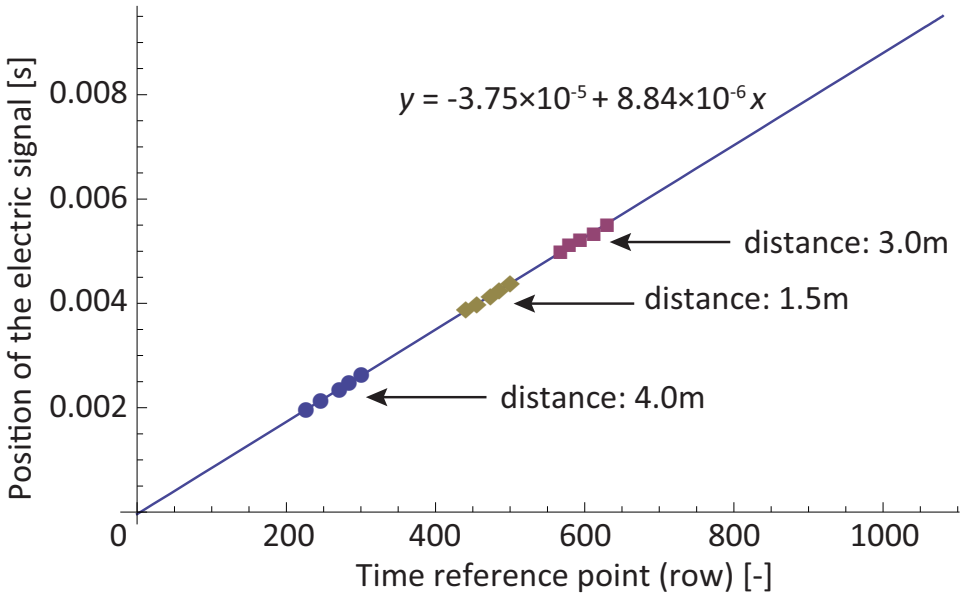


Figure 4.9. Correspondence of acquired time to electric signal.

As the slope of the line is 8.84×10^{-6},

$$1080 \cdot 8.84 \times 10^{-6} = 0.00954\,\text{s} \tag{4.7}$$

is the scanning time across the frame. The frame interval is 0.0167 s, and hence the percentage of dead time is:

$$100 \cdot (1 - 0.00954/0.0167) = 42.8\%. \tag{4.8}$$

4.3.3. Summary

We have introduced a basic idea for time synchronization using a CMOS image sensor. Its feasibility has been confirmed by experiment. First, we could obtain the time reference from the video image that captured the modulated LED light. Then, by comparing with an electric signal, the time reference was proved to be consistent in the video frame. We could then perform ToA measurements with acoustic waves using the time reference provided by the video images.

4.4. 3D Localization Experiment

In the last section, we confirmed the feasibility of the time synchronization method using LED light. Here, we conduct 3D ToA localization experiments with time synchronization. We name this this method *SyncSync* because it uses sync patterns for acoustic bursts and light synchronization. We also perform 3D TDoA localization to compare the difference between ToA and TDoA.

4.4.1. Experimental Setup

Figure 4.10 shows a diagram of the experimental system. We used an LED floodlight consisting of 56 white LEDs (OSW54L5111P, OptoSupply). We tried using infrared light as we did in the last section, however, we could not achieve synchronization because the smartphone we used did not capture infrared light.

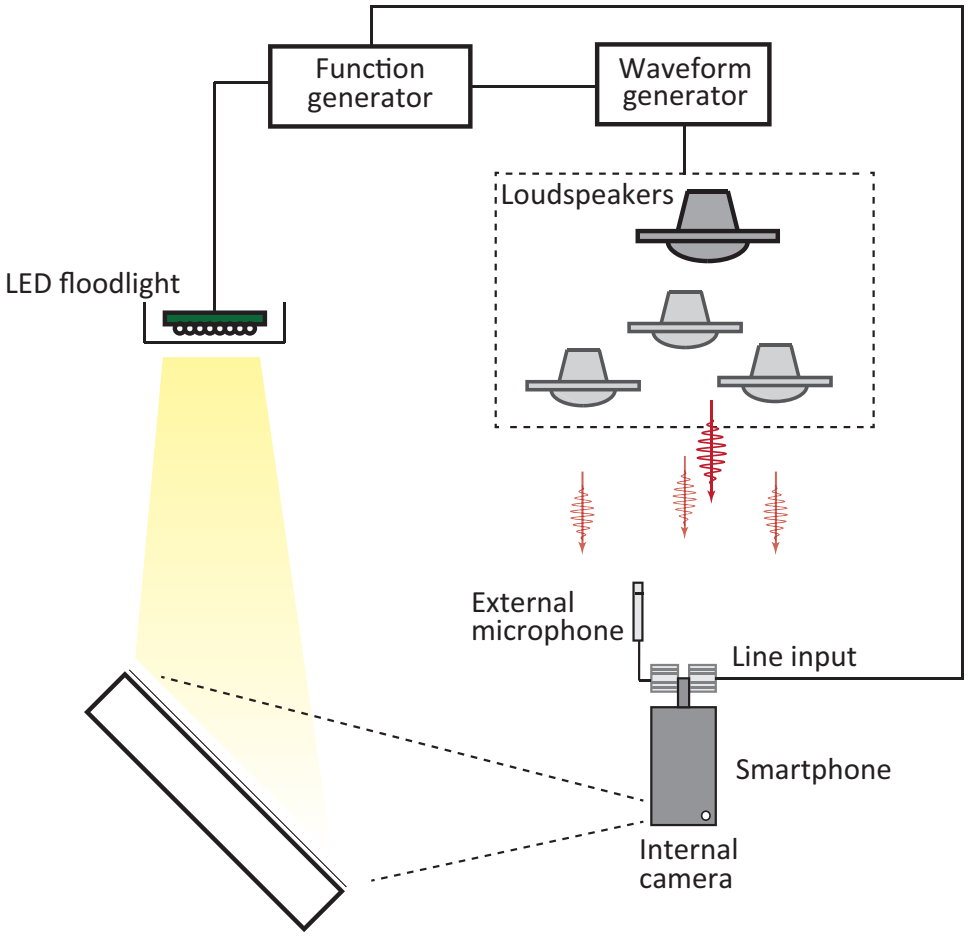
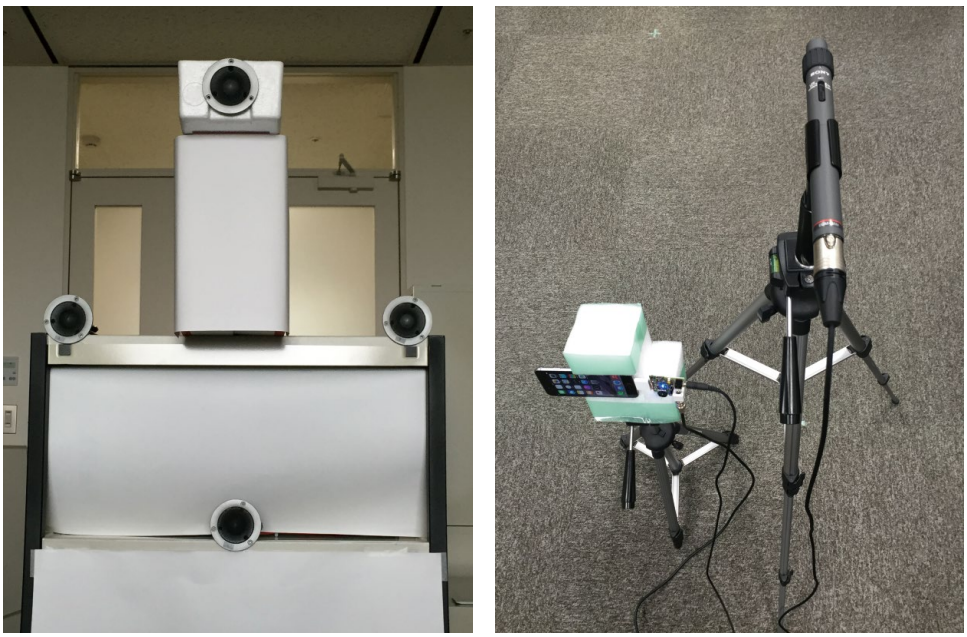


Figure 4.10. Experimental system for 3D localization.

The floodlight was covered by a translucent plastic film to diffuse rays emitted directly from the LED elements. The light was modulated with a 30.0 Hz square wave generated by a function generator (AFG3102, Tektronix) and was used to illuminate wallpaper on the back of a cabinet. The function generator was connected also to the smartphone's line input so that this signal could be used to validate the time reference found by frame analysis and to perform acoustic localization for evaluating precision. A lab-built multichannel waveform generator produced sound waves that were emitted by an array of tweeters (FT200D, FOSTEX; Figure 4.11(a)) triggered by the function generator's timings.

A smartphone (iPhone 6, Apple) with an external stereo condenser microphone (iM2, TASCAM) was used as the target smartphone in this study (Figure 4.11(b)). Both channels of the microphone elements were removed (Figure 4.12). The left channel element was replaced with the line input of the smartphone, and the right channel element was used as the interface for the reference microphone (ECM-G5M, SONY) input. The smartphone's rear camera took

video images of the wallpaper lit by the floodlight. The motion image had a 1920×1080 pixel format and a 60.0 fps (progressive) frame rate. The audio sampling rate was 48.0 ksps. Captured video images were stored as MPEG-4 files that were transferred to a PC and split for separate analysis into image and audio files using the FFmpeg tool (http://www.ffmpeg.org/).



(a) Tweeters (loudspeakers) (b) Smartphone with a microphone

Figure 4.11. Tweeters and smartphone.

Figure 4.13 shows the square wave modulating the LED floodlight and the audio signals. The reference acoustic pulses sent to the smartphone's line input were chosen in sync pattern forms, so they could be analyzed to obtain precise sub-microsecond time resolutions. The reference timing was chosen according to the rising edge of the square wave driving the LEDs. Timing waveforms were generated once every three LED cycles. Ten carrier frequencies between 10.75 and 15.25 kHz were chosen, and five sync patterns—one reference signal to the smartphone's line input and the others as airborne sound waves—were formed using adjoining carriers. The frequency separation in each sync pattern was 500 Hz, so the duration of two-beat sync patterns was $1/500 \times 2 \times 1000 = 4.0$ ms.

Experiments were performed in an office room with the ceiling lights off. Four loudspeakers (anchor nodes) were placed at the coordinates listed in Table 4.1. The largest distance between any two of the four loudspeakers was 0.618 m. The microphone was placed at one of the targets shown in the same table and illustrated in Figure 4.14. There were no obstacles between the devices. We performed 150 measurements from each positioning spot. Localization processes were performed using a mathematical computation program (Mathematica, Wolfram).

Setting the peak driving power of each loudspeaker to 0.14 W caused them to generate a sound pressure of approximately 78 dB at a distance of 1 m. The LED floodlight was placed 0.8 m above the floor with 0.6 m separation from the reflector wallpaper. The peak LED driving power was 9.2 W. The video capture duration was 15 s for each measurement.

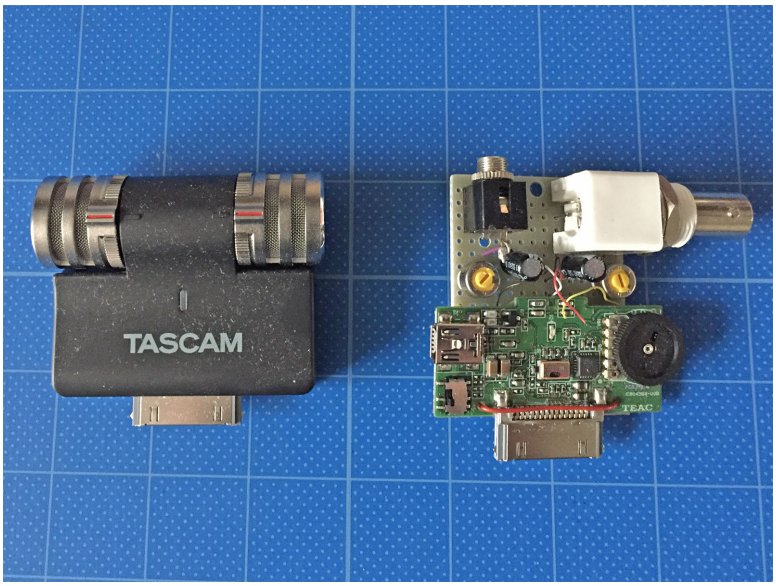


Figure 4.12. External microphones. Original (left) and modified (right).

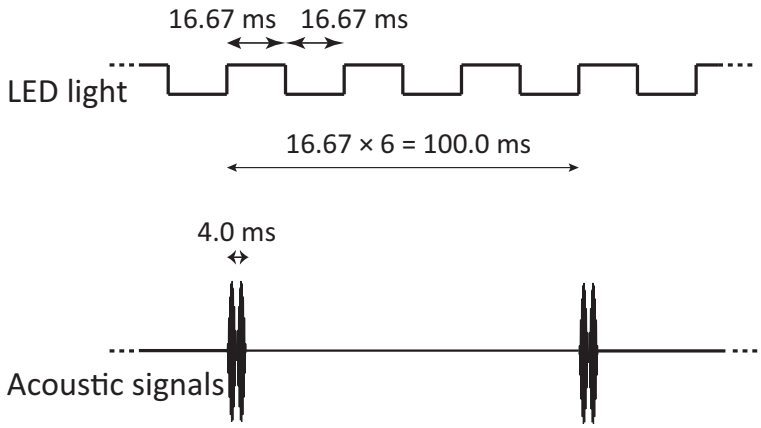


Figure 4.13. LED modulation pattern and audio signal.

4.4.2. Results

Precision of Time Synchronization

The acquired data were analyzed as follows. We coupled adjoining image frames that corresponded to an A-shaped frame and a V-shaped frame. By averaging the pixel values in each horizontal row, we created the vectors $A(x)$ and $B(x)$ for the A- and V-shaped frames, respectively. When the background luminance distribution $b(x)$ was set at the minimum pixel value of $B(x)$, we could solve equation (4.1) to determine $p(x)$ and $f(x)$. Consequently, we obtained the time reference point x at which $p(x)$ had a maximum value. The time reference points calculated by the frame pairs fluctuated, but their standard error was never more than 0.5 rows. Figure 4.15 shows the standard error of the estimated time reference points against the number of row pairs used to calculate the error. This shows that within 1 s we obtained a time synchronization precision of 0.03 rows, which is equivalent to 0.17 mm of an airborne acoustic wave. This is precise enough for the time reference of ToA measurements.

Table 4.1. Coordinates of loudspeakers and microphone positions.

Device	Coordinates (x, y, z) (m)
Loudspeaker #1	(0.238, 1.815, 1.685)
Loudspeaker #2	(-0.005, 2.000, 1.395)
Loudspeaker #3	(0.502, 2.000, 1.395)
Loudspeaker #4	(0.245, 2.000, 1.095)
Microphone position #1	(0.750, 0.000, 0.977)
Microphone position #2	(0.500, 0.000, 0.977)
Microphone position #3	(0.250, 0.000, 0.977)
Microphone position #4	(0.000, 0.000, 0.977)
Microphone position #5	(-0.250, 0.000, 0.977)
Microphone position #6	(0.750, 0.520, 0.977)
Microphone position #7	(0.500, 0.520, 0.977)
Microphone position #8	(0.250, 0.520, 0.977)
Microphone position #9	(0.000, 0.520, 0.977)
Microphone position #10	(-0.250, 0.520, 0.977)

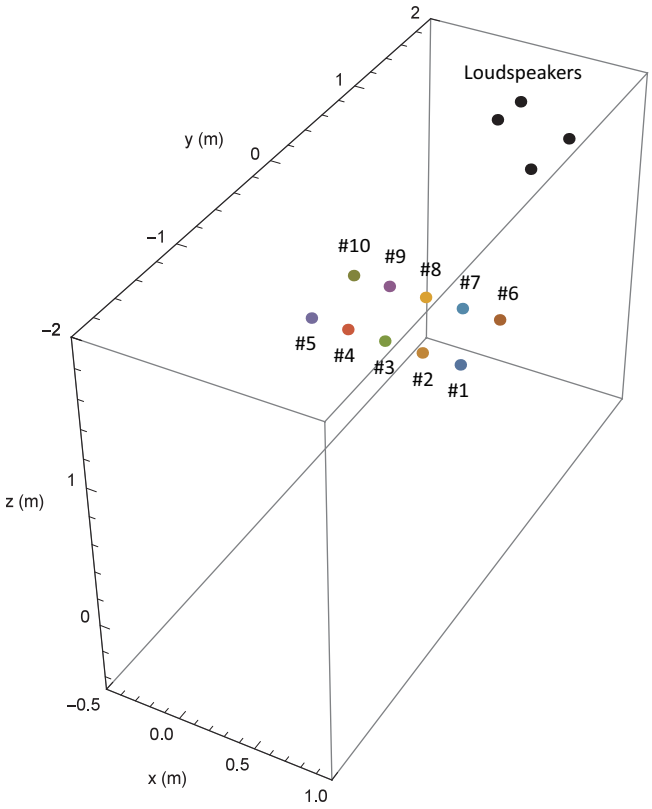


Figure 4.14. Loudspeakers and true positions.

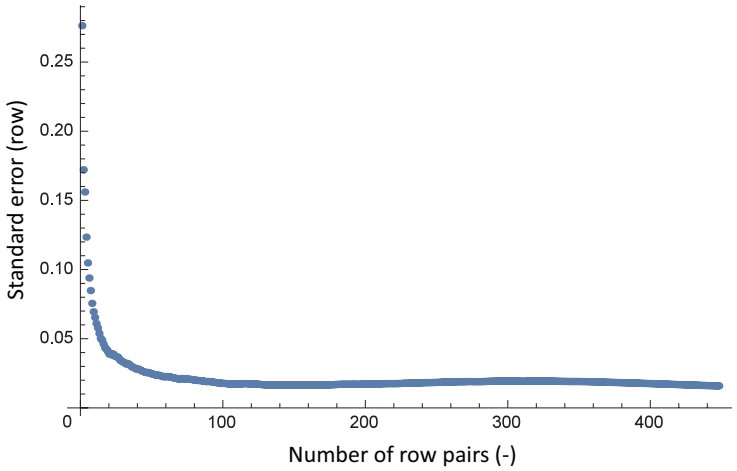


Figure 4.15. Stability of time reference acquired from motion image.

4.4.3. Comparison between ToA and TDoA

The audio file contains the electric reference signal produced by the function generator and the acoustic signal picked up by the microphone. Using FDM-PAM, all the epochs in each signal were determined. The standard deviation of the position timing was 0.26 µs for the electric signal and 0.88 µs for the acoustic one. The precision of this acoustic signal is equivalent to a distance of 0.36 mm. This implies that we can deploy accurate multilateration with a short baseline setting.

Table 4.2 shows the statistical results of the localization, and Figs. 4.16 and 4.17 illustrate the distribution of the results. In TDoA, the area of the distribution was elongated in the loudspeaker–microphone direction, and there were offset errors that have no relation to the true positions. In ToA, the obtained results were closer to the true positions; the standard deviations were 10 mm or less. Furthermore, the offsets were smaller and were consistent with the true positions, so they can be reduced by advance surveying.

As mentioned above, offset errors were observed in both TDoA and ToA measurements; they are mainly due to the radiation characteristics of the loudspeakers and multipath signals. Measurements were also affected by reflections from the neighboring walls, especially at the measurement sites at each side. Although these disturbances are common for both TDoA and ToA measurements, the localized positions are different between the measurements because of the difference in localization schemes. As we described in Subsection 4.2.4, the smartphone must be positioned at one of the intersections of three surfaces. In TDoA, with a short baseline setting, the surfaces of any two hyperboloids intersect each other at angles smaller than those of ToA, and the intersection curve follows the loudspeaker–microphone direction. Therefore, if the uninvolved surface *moves* due to a slight measurement error, then the intersection of the three surfaces shifts a large distance along the loudspeaker–microphone direction. In ToA, the surfaces of spheres intersect at larger angles, and therefore the effect of a measurement error is smaller than in TDoA.

In the processes of solving equation (4.2) or (4.4), the residuals ($\sum \| \cdot \|^2$) were nearly zero in TDoA #2–#6, TDoA #10, and ToA #1–#10. In these cases, the surfaces determined from equation (4.3) or (4.5) had intersections substantially at the localized positions. In contrast, in TDoA #1 and TDoA #7–#9, some residuals were found. In these cases, the surfaces determined

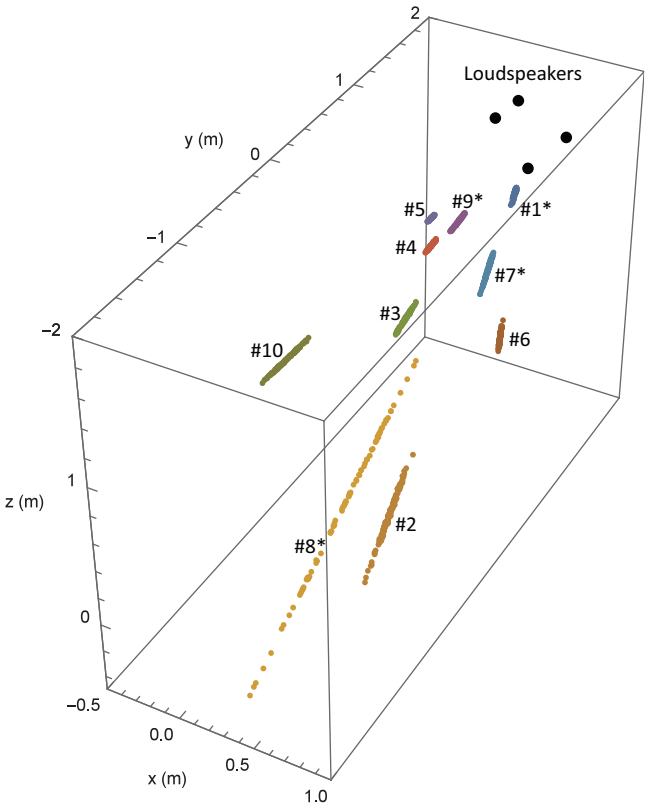


Figure 4.16. TDoA localization results (* denotes approximations).

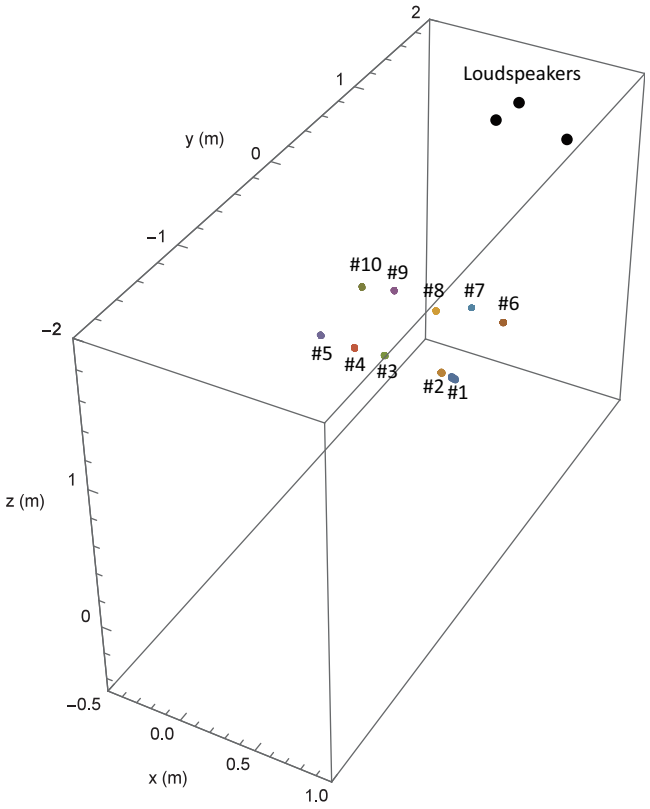


Figure 4.17. ToA localization results.

Table 4.2. 3D localization results (* denotes approximations).

Localization	Mean (m)	Standard deviation (m)		Offset
	(x, y, z)	(x, y, z)	Overall	(m)
TDoA #1*	(0.376, 1.514, 1.204)	(0.007, 0.028, 0.008)	0.030	1.576
TDoA #2	(0.797,-0.937, 0.549)	(0.028, 0.153, 0.046)	0.162	1.073
TDoA #3	(0.258, 0.270, 0.974)	(0.002, 0.056, 0.015)	0.058	0.270
TDoA #4	(0.132, 0.839, 1.125)	(0.003, 0.029, 0.008)	0.030	0.863
TDoA #5	(0.036, 1.024, 1.194)	(0.003, 0.014, 0.004)	0.015	1.085
TDoA #6	(0.831, 0.392, 0.954)	(0.014, 0.039, 0.012)	0.043	0.153
TDoA #7*	(0.505, 0.880, 1.040)	(0.016, 0.070, 0.020)	0.075	0.365
TDoA #8*	(0.486,-2.127,-0.163)	(0.215, 3.738, 1.400)	3.997	2.891
TDoA #9*	(0.146, 1.191, 1.112)	(0.004, 0.037, 0.011)	0.039	0.700
TDoA #10	(-0.322,-0.182, 0.751)	(0.022, 0.081, 0.026)	0.087	0.741
ToA #1	(0.728,-0.031, 0.900)	(0.006, 0.004, 0.007)	0.010	0.086
ToA #2	(0.615, 0.037, 0.858)	(0.002, 0.001, 0.003)	0.004	0.170
ToA #3	(0.260,-0.013, 0.892)	(0.002, 0.001, 0.002)	0.003	0.087
ToA #4	(0.049,-0.001, 0.868)	(0.002, 0.001, 0.002)	0.003	0.120
ToA #5	(-0.183,-0.007, 0.886)	(0.002, 0.001, 0.002)	0.003	0.113
ToA #6	(0.792, 0.502, 0.991)	(0.001, 0.001, 0.001)	0.002	0.047
ToA #7	(0.587, 0.487, 1.039)	(0.001, 0.001, 0.001)	0.002	0.112
ToA #8	(0.334, 0.510, 0.919)	(0.001, 0.000, 0.001)	0.001	0.103
ToA #9	(0.065, 0.472, 1.004)	(0.001, 0.001, 0.001)	0.002	0.085
ToA #10	(-0.155, 0.459, 0.967)	(0.001, 0.000, 0.001)	0.002	0.113

from equation (4.5) had no intersections. The equations had no real number solutions, therefore those results were approximations. For TDoA #8, in particular, the calculations of equation (4.4) were unstable. TDoA calculation is considered inherently unstable when some measurement errors are present. Judging from the residual of each measurement in TDoA #8, 17 out of 150 measurements were excluded as outliers in Table 4.2 and Figure 4.16; despite this, the distribution is still wide because of the instability of the calculation. In practical applications we should eliminate the localized positions from some area that cannot receive acoustic signals, but we have shown all localized positions (except outliers) in Table 4.2 and Figs. 4.16 and 4.17 to compare the characteristics of TDoA and ToA calculation.

4.4.4. Evaluation using DOP

Here, we evaluate the experimental results by using DOP [27]. Although DOP has been examined in detail in the GNSS field including the GPS [22], it has been also applied to indoor localization [39]. The DOP values are found from linearized forms of equations that are solved iteratively. In this study, we used Mathematica to solve ToA and TDoA equations, and did not become involved in the detailed calculation processes. Therefore, the underlying equations for Mathematica's solver might be different from those of DOP. However, because DOP is determined from only the geometrical arrangement of loudspeakers and a smartphone,

we attempted to apply it for error evaluation in our experiments.

Position dilution of precision (PDOP) is one of the DOP parameters and is defined as

$$\mathrm{PDOP} = \frac{\sqrt{{\sigma_x}^2 + {\sigma_y}^2 + {\sigma_z}^2}}{\sigma_0}, \tag{4.9}$$

where ${\sigma_x}^2$, ${\sigma_y}^2$, and ${\sigma_z}^2$ are the predicted variances of localization (when the input signal has a unit variance) in the x-, y-, and z-directions, respectively; and ${\sigma_0}^2$ is the variance of the range measurement (see Appendix B for the derivation). Values $\sigma_0\sigma_x$, $\sigma_0\sigma_y$, and $\sigma_0\sigma_z$ are the predicted standard deviations in x-, y-, and z-directions, respectively; σ_0PDOP is the predicted overall standard deviation.

We show DOP values in Table 4.3. We used the mean values of the localized smartphone positions, not the true positions, to allow evaluating the experimental results. The PDOP values in TDoA are worse than those of ToA. In TDoA, σ_y is the largest among σ_x, σ_y, and σ_z and is close to PDOP (i.e., σ_y is the principal factor for PDOP). In contrast, in ToA, σ_y is smaller than σ_x or σ_z.

In this study, σ_0 in each measurement has a value in the range 0.255–0.597 mm, which was derived from the intervals of received sound waves. To validate the localization errors in Table 4.2, we calculate the ratios of localization errors to predicted errors, as also shown in Table 4.3. In TDoA #2–#6, TDoA #10 and ToA #1–#10, the ratios are close to unity, which means that DOP values represent error distribution characteristics well. In contrast, in TDoA #1 and TDoA #7–#9, the ratios are larger than those of the other cases are, but we can see the tendency of the distribution. Note that TDoA #1 and TDoA #7–#9 are the cases in which the localization calculation had some residuals (the simultaneous equations had no real number solutions); we assume these residuals might be the reason for the larger ratios. In the case of TDoA #8, the ratios are much larger than in the other cases. We consider that this is caused by the instability in solving ill-conditioned equations. In this situation, the localization distribution becomes quite wide, and DOP cannot correctly handle the wide distribution because it is derived from the first-order approximation of an equation and is better suited for cases with only small deviation (see Appendix B). Finally, despite having ignored errors in the installation positions of the loudspeakers, which could affect the DOP values, the tendency of the distribution was observed clearly.

4.4.5. Alignment of Video and Audio Frames

In this study, the acquired MPEG-4 files have timestamps in their video and audio frames, but the resolution is only 1/600 s. Because this is too coarse for time synchronization, we used the function generator's reference signal for a more precise evaluation. Smartphone applications may be able to refer to the timestamps in video and audio buffers in the OS. We analyzed the variation of frame intervals of the video and audio buffers (iPhone 6s Plus, Apple), and the standard deviations were 1.94 µs for the video buffer and 0.47 ns for the audio one. These values are sufficient for aligning video and audio frames, therefore we can map the time reference points in the video track onto the corresponding points in the audio track.

Consequently, the function generator's reference signal will not be needed for the localization system. As a result, the wired connection between the function generator and the smartphone

Table 4.3. DOP Values and Ratios of Measured Standard Deviation to Estimated Standard Deviation (* denotes figures are based on approximations).

Localization	DOP Values (-)				Ratio (-)			
	$(\sigma_x, \sigma_y, \sigma_z)$			PDOP	(x, y, z)			Overall
TDoA #1*	(3.1,	10.9,	4.7)	12.3	(3.92,	4.31,	2.71)	4.09
TDoA #2	(89.0,	472.8,	140.5)	501.2	(0.87,	0.88,	0.89)	0.88
TDoA #3	(5.1,	155.2,	40.1)	160.4	(0.81,	0.99,	1.04)	0.99
TDoA #4	(7.1,	63.8,	16.4)	66.2	(1.36,	1.28,	1.38)	1.28
TDoA #5	(9.4,	42.5,	10.2)	44.7	(0.95,	0.93,	1.02)	0.94
TDoA #6	(43.8,	119.8,	35.0)	132.3	(0.95,	0.95,	1.00)	0.95
TDoA #7*	(13.7,	58.6,	20.1)	63.5	(3.89,	4.09,	3.42)	4.02
TDoA #8*	(62.9,	1065.,	410.3)	1143.	(3.40,	13.77,	13.39)	13.72
TDoA #9*	(4.4,	30.2,	11.6)	32.7	(3.35,	4.44,	3.56)	4.32
TDoA #10	(64.4,	245.1,	75.4)	264.3	(1.10,	1.05,	1.08)	1.05
ToA #1	(6.0,	2.6,	7.5)	10.0	(1.66,	2.46,	1.52)	1.65
ToA #2	(5.8,	2.6,	7.1)	9.5	(1.08,	1.07,	1.00)	1.04
ToA #3	(5.8,	2.3,	7.1)	9.5	(0.86,	0.80,	0.68)	0.76
ToA #4	(5.8,	2.4,	7.1)	9.5	(1.00,	0.77,	0.80)	0.88
ToA #5	(5.9,	2.6,	7.3)	9.8	(0.85,	0.67,	0.88)	0.85
ToA #6	(4.6,	2.4,	5.7)	7.7	(0.86,	0.83,	0.72)	0.79
ToA #7	(4.5,	2.0,	5.5)	7.4	(0.75,	0.94,	0.83)	0.81
ToA #8	(4.4,	2.2,	5.2)	7.2	(0.85,	0.74,	0.63)	0.73
ToA #9	(4.5,	1.9,	5.4)	7.3	(0.89,	0.97,	0.97)	0.94
ToA #10	(4.7,	2.3,	5.6)	7.6	(0.96,	0.63,	0.82)	0.86

will be removed. Then, a localization application using the SyncSync method works as a standalone system.

4.5. Summary

We have confirmed the feasibility of ToA localization using the SyncSync method. Although the experiments showed some systematic errors, which could be compensated for by calibration, random errors were 10 mm or less. This was accomplished by the accurate time synchronization mechanism of SyncSync. The beacon arrangement in this study had a very *short* baseline. Furthermore, ToA is easier to set up because it requires fewer beacons and has a shorter baseline configuration. We also compared the experimental results with the estimated errors from DOP. The values were nearly the same: DOP showed good estimation of the measurements.

5. Smartphone Localization Application with Visible Light Communication

5.1. Introduction

In Chapter 4, we introduced an indoor 3D localization system for smartphones. We used the smartphone's microphone to detect acoustic signals and its video camera for time synchronization between sender nodes and the smartphone. We conducted ToA measurements based on the time synchronization, and this offered better performance than did TDoA. In our ToA measurements, the standard deviations were less than 10 mm, whereas for TDoA they were 10 mm to 100 mm. In the worst cases, the positioning calculations diverged.

In this chapter, we describe our new time synchronization technique based on VLC. Our experimental results show that the performance of VLC-based synchronization is comparable to that of dedicated systems. As VLC becomes more prevalent for mobile devices, our method will become useful for both communication and localization.

Many of the parameters that are needed for localization, such as transmitter coordinates and the frequencies used for acoustic waves, can be conveyed by VLC. Hence, the application software does not need to refer to an external database for this information, and the system can work without network connectivity. Furthermore, user privacy is protected because no individual user is announced to others. These are some of the advantages of the proposed method.

Our aim is to realize localization with this scheme. First, we verified the concept of the proposed method and found some issues with it. Then, we made a localization application and evaluated the performance. We discuss some issues concerning this method.

5.2. System Overview

We aim to realize ToA-based acoustic localization that uses VLC for time synchronization. We use a transmitter module that consists of loudspeakers and an LED. The loudspeakers emit short bursts of sound simultaneously, which are registered by a smartphone. The smartphone then calculates the ToF of each burst, and determines its position using ToA-based multilateration. The LED is modulated for VLC, the carrier wave of which is synchronized to the emitted bursts. In the demodulation process, the VLC carrier signal is extracted and used to estimate the emission time; this is the time synchronization of the system. The ToF is the time between burst emission and reception.

5.2.1. Visible Light Communication

In Chapter 4, we modulated an LED with a square wave whose frequency was exactly half the frame rate of the video camera. Capturing the modulated light by a CMOS image sensor, we observed a gradation pattern in the obtained image because of the rolling shutter effect. By assuming that the exposure time was the video-frame period, the profile of the pattern in the shutter-scanning direction becomes part of a triangle wave. We detected either the top or bottom of the profile and used it as a reference point for time synchronization.

We now consider the case in which the modulation frequency of the square wave is one third of the video frame rate (Figure 5.1 (a)). The interval of integration of the square wave is the exposure time of the video camera, and the rolling shutter works as a sliding window function. Applying the window function for three consecutive frames (i.e., the period of the LED modulation), we obtain the pixel value profile shown in Figure 5.1 (b). This can be approximated as a sinusoidal wave whose phase denotes when the LED is turned on or off.

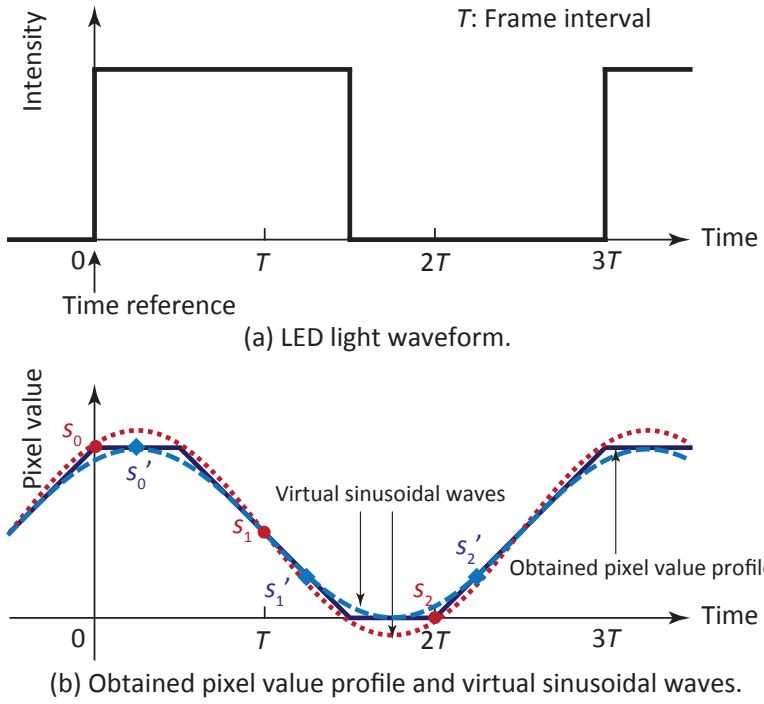


Figure 5.1. LED light waveform, obtained pixel value profile, and virtual sinusoidal waves.

5.2.2. Three-point Demodulation Method

We use the smartphone camera to capture a video of the illuminated LED. In each frame, the group of pixels that contains the LED is regarded as the region of interest (RoI). The *triad* $\boldsymbol{s} = (s_0, s_1, s_2)^{\mathrm{T}}$ is the set of mean RoI pixel values for three consecutive frames. It is used to detect a complex sinusoidal wave

$$f(\theta) = Ae^{j\theta} + Be^{-j\theta} + C, \tag{5.1}$$

where θ denotes the phase of the sinusoidal wave, A and B express its amplitude, C is the constant component of the wave, and $j = \sqrt{-1}$. Note that the unknown values of A, B, and C

are complex. Assigning $\theta = 0,\ 2/3\pi$ and $4/3\pi$ in (5.1), we get a matrix equation

$$\begin{pmatrix} s_0 \\ s_1 \\ s_2 \end{pmatrix} = \begin{pmatrix} f(0) \\ f\left(\frac{2}{3}\pi\right) \\ f\left(\frac{4}{3}\pi\right) \end{pmatrix} = \begin{pmatrix} 1 & 1 & 1 \\ \omega & \omega^2 & 1 \\ \omega^2 & \omega & 1 \end{pmatrix} \begin{pmatrix} A \\ B \\ C \end{pmatrix} = M \begin{pmatrix} A \\ B \\ C \end{pmatrix}, \tag{5.2}$$

where

$$\begin{aligned} \omega &= \frac{-1 + j\sqrt{3}}{2}, \\ \omega^2 = \bar{\omega} &= \frac{-1 - j\sqrt{3}}{2}, \\ M &= \begin{pmatrix} 1 & 1 & 1 \\ \omega & \omega^2 & 1 \\ \omega^2 & \omega & 1 \end{pmatrix}. \end{aligned} \tag{5.3}$$

Here, $\bar{\cdot}$ denotes the complex conjugate, and ω, ω^2, and $\omega^3 = 1$ are the three cube roots of 1. We can solve (5.2) as

$$\begin{pmatrix} A \\ B \\ C \end{pmatrix} = \frac{1}{3} \begin{pmatrix} 1 & \omega^2 & \omega \\ 1 & \omega & \omega^2 \\ 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix} = \frac{1}{3} M^* \begin{pmatrix} s_1 \\ s_2 \\ s_3 \end{pmatrix}. \tag{5.4}$$

Where

$$M^* = \begin{pmatrix} 1 & \omega^2 & \omega \\ 1 & \omega & \omega^2 \\ 1 & 1 & 1 \end{pmatrix}. \tag{5.5}$$

is the conjugate transpose of M.

Because $\boldsymbol{s}$ is a vector whose components are real, $B = \bar{A}$ holds. Then, (5.1) can be written as

$$f(\theta) = a\cos(\theta + b) + c, \tag{5.6}$$

where the unknowns a, b, and c are real numbers that satisfy

$$\begin{aligned} a &= 2|A|, \\ b &= \arg A, \\ c &= C. \end{aligned} \tag{5.7}$$

The sinusoidal function (5.6) is not the original square wave that modulated the LED, but it corresponds to it uniquely. Therefore, we refer to (5.6) as the *virtual sinusoidal wave*.

We set the time reference as the rising edge of the original square wave. Because the LED modulation and the exposure time of the video camera are not synchronized, the time at which the RoI pixel values are sampled varies in each measurement. Two examples of triads, $\boldsymbol{s} = (s_0, s_1, s_2)^{\mathrm{T}}$ and $\boldsymbol{s'} = (s'_0, s'_1, s'_2)^{\mathrm{T}}$, are shown in Figure 5.1 (b). Although each triad yields a different sinusoidal wave, we can determine the time reference by using (5.6) in each case.

5.2.3. Converting Virtual Sinusoidal Wave to the Original Square Wave

Let us consider that the points on the pixel value profile move along the profile. Suppose that s_0, s_1, and s_2 of $\boldsymbol{s}$ in Figure 5.1 (b)) move from $t = 0$ to $t = 3T$, from $t = T$ to $t = 4T$, and from $t = 2T$ to $t = 5T$, respectively. Then, plotting the amplitude and the initial phase derived from the triad (s_0, s_1, s_2) in polar coordinates, we can get a hexagon. After getting the phase and the amplitude of the virtual sinusoidal wave, they are to be converted to those of the original square wave.

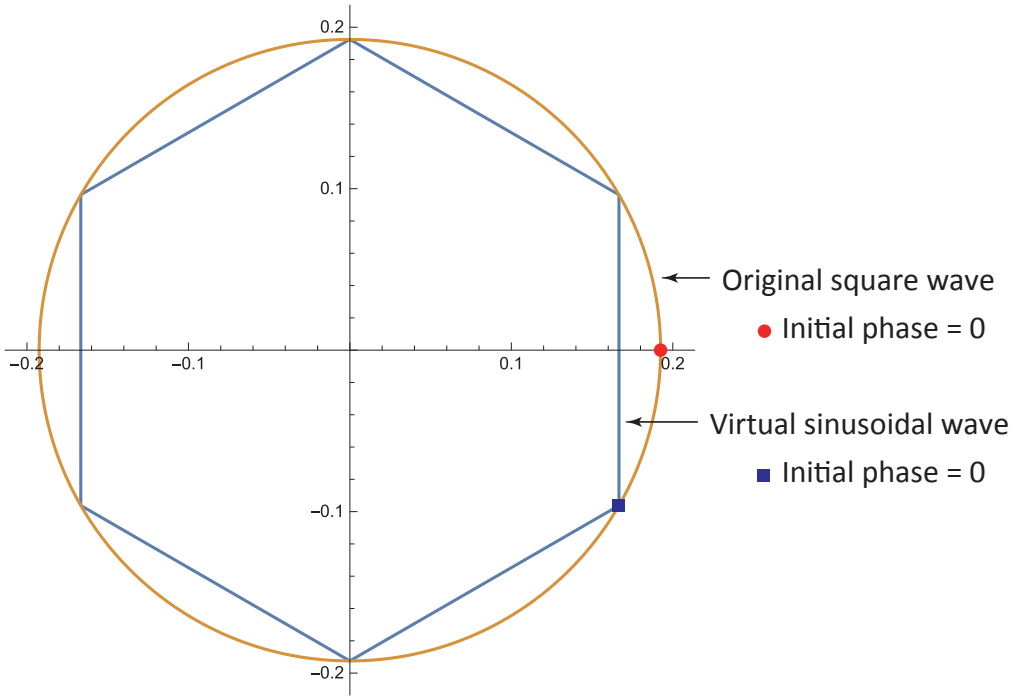


Figure 5.2. Polar coordinates plot for the virtual sinusoidal wave and the original square wave.

5.2.4. Extracting Time Reference from VLC Messages

We modulate the LED with six square wave patterns. They have the same frequency, amplitude, and mean, but their initial phases are separated by $2/3\pi$. This is a phase-shift keying (PSK) scheme that uses six symbols. The constellation diagram is shown in Figure 5.3.

As explained above, a pixel value is the result of integrating the LED output during an exposure. If the exposure time spans the boundary of the LED's lighting pattern, the integration of this term is not suitable for demodulation. To avoid this situation, we introduce a *guard frame* that is similar to the guard interval [34] used in telecommunications. Appended to the original frame, a guard frame is the same lighting pattern as the first frame of the original pattern. Accordingly, the signal length of a symbol becomes four times that of the frame duration. This method ensures that we can select three consecutive frames from a video sequence that have the correct integral values of the original waveform. Figure 5.4 shows the waveform for symbol "5", whose initial phase is $-1/3\ \pi$.

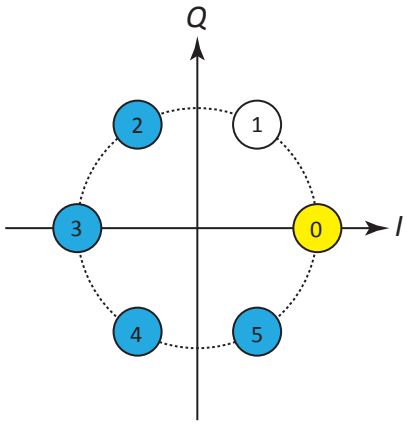


Figure 5.3. Constellation diagram for 6-PSK.

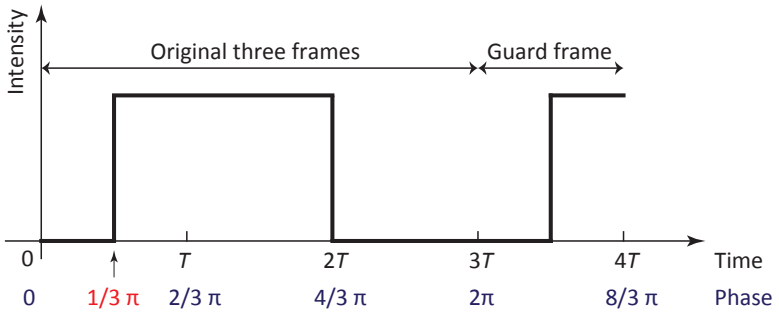


Figure 5.4. Waveform of symbol “5”.

5.2.5. Timestamps of Video and Audio Frames

To calculate the precise ToF, we need to know the relationship between the video frame and the audio frame. As we mentioned in 4.4.5, the resolution of the timestamps in the MPEG-4 files was only 1/600 s. then, we used the timestamps of the video and audio buffers using OS’s CMSampleBufferGetPresentationTimeStamp function, which have 1 ns resolution.

First, we measured the intervals of the video and audio frames. For video buffers, the mean and the standard deviation were 16.16 ms and 1.94 µs when we set the video frame rate as 60 fps. For audio buffers, the mean and the standard deviation were 85.33 ms and 0.47 ns when we set the audio sampling rate as 48ksps, in this case the buffer size was 4096 bytes. For both cases, the means are equal to the theoretical values, and the standard deviations are enough small for time synchronization and ToF measurement.

We once tried to use the OS’s gettimeofday function, which has 1 µs resolution, at the beginning of the callback function that is called when the buffer is filled with data. However, the precision was not sufficient. The standard deviation of the gettimeofday’s value compared with that of CMSampleBufferGetPresentationTimeStamp was 2.6 ms, which is not suitable for precise ToF measurement.

5.3. Time Synchronization Experiment using VLC

5.3.1. Experimental Setup

We conducted an experiment to evaluate the feasibility of this method. Figure 5.5 shows a diagram of the experimental system. In the transmitter, an LED floodlight is placed behind a circular translucent window (20 cm in diameter) to emit the modulated light (Figure 5.6). The LED floodlight consisted of 56 red LEDs (OS5RKA5111A, OptoSupply; dominant wavelength: 624 nm). The waveform signal of the LED modulation was generated by an arbitrary function generator (AFG3102, Tektronix). This signal drove a lab-built voltage-to-current converter that supplied electric current to the LED floodlight.

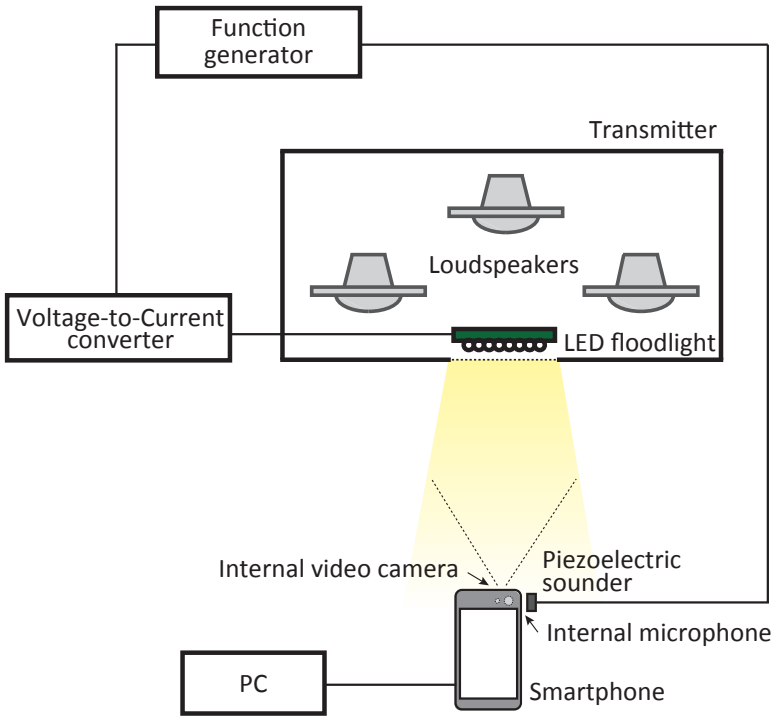


Figure 5.5. Experimental system for VLC time synchronization.

Three loudspeakers were mounted on the transmitter (Figure 5.6) for acoustic localization, but they were not used in this experiment. Instead, a piezoelectric sounder (PKM13EPYH4000-A0, Murata Manufacturing) was attached directly to the smartphone microphone to evaluate the performance of the smartphone's audio signal processing. Acoustic waves were emitted in 4 ms bursts through the sounder at the beginning of each LED modulation cycle. PAM was used to determine the precise time of signal reception.

A smartphone (iPhone 6s Plus, Apple) was mounted on a tripod at a distance of 1.0 m, 1.5 m, or 2.0 m from the LED to capture video images. We developed a video-capture application to allow the smartphone to record the pixel values of two RoIs (RoI1 and RoI2). Because they were separated in the shutter-scanning direction, the detected phases of RoI1 and RoI2 were different. The frame rate of the video camera was 60 fps. Although demodulation could be processed with the video-capture application, the recorded pixel values were sent as logging messages and analyzed offline by a PC to evaluate the performance statistically. Timestamps

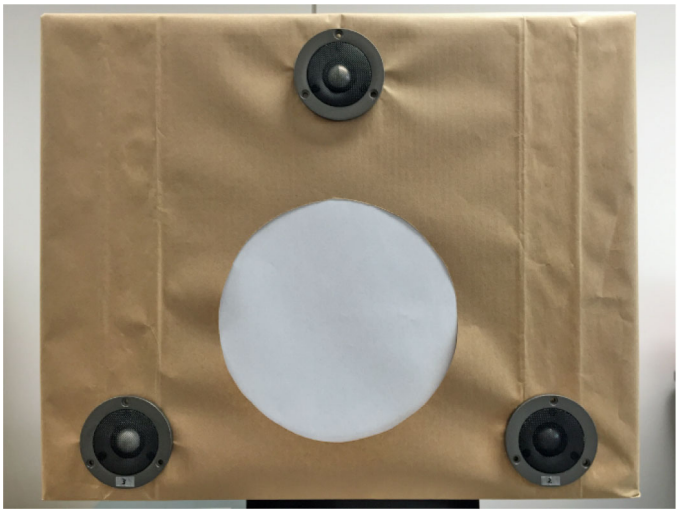


Figure 5.6. Transmitter for VLC time synchronization.

of the video frames were extracted from the sampling buffer of the smartphone through its application programming interface.

Audio signals captured by the smartphone's microphone were recorded alongside the video images as MPEG-4 files that were transferred to the PC. The audio sampling rate was 48.0 ksps. Timestamps of the audio buffer were also extracted and recorded for analysis.

The LED's modulating signal was generated as follows. The signal consisted of a preamble and a payload. The preamble was used to detect the existence of a signal and to find the top frame of a triad. It comprised 12 "0" symbols of the 6-PSK. For the payload, we sent the text message "HELLO, WORLD". Each character in the message was converted to a 6-bit binary code and divided into three 2-bit pieces that were assigned to the symbols "2"–"5". The symbol "1" was not used for data transmission but its existence was used as an integrity check on the receiver side. The payload of the transmission had 36 symbols, meaning that the total length of the signal was 48 symbols. As each symbol occupied four video frames, the duration of the signal was 3.2 s. The signal was transmitted repeatedly, and the video capture duration was 30 s for each measurement.

5.3.2. Results

Performance of VLC Demodulation and Time Synchronization

The pixel values captured by the smartphone application were demodulated on a PC. We assume a sequence of pixel values $p_i (i = 1, 2, 3, \dots)$ in an RoI. We can make a series of triads as $t_1 = (p_1, p_2, p_3), t_2 = (p_2, p_3, p_4), t_3 = (p_3, p_4, p_5), t_4 = (p_4, p_5, p_6), \dots$. As mentioned before, there is a *right* triad that does not span two symbols in $t_1 \dots t_4$. A right triad appears every four triads. Triads in the preamble are used to select the correct sequence of triads in the payload.

A typical constellation diagram of the results is shown in Figure 5.7. Here, the plotting plane is rotated so that the mean phase of RoI1's preamble becomes zero. The symbols are definitely separated, which means that the message was perfectly demodulated. The phase difference between RoI1 and RoI2 is due to the scanning speed of the rolling shutter. Using this speed and the phase of RoI1 or RoI2, we can estimate the phase of the first scanning line of the image

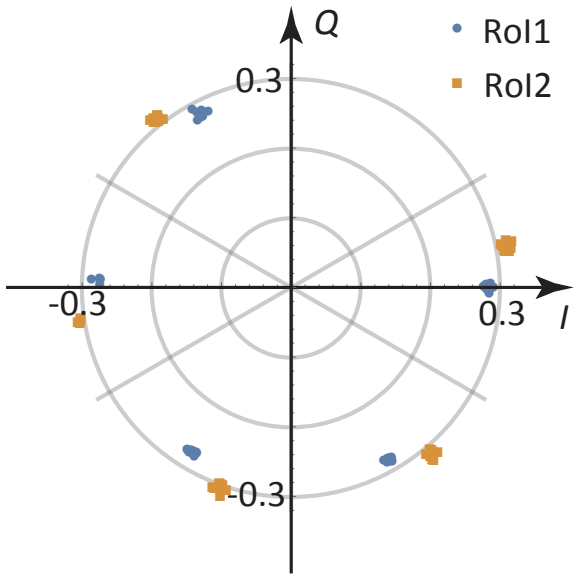


Figure 5.7. Constellation diagram of the results.

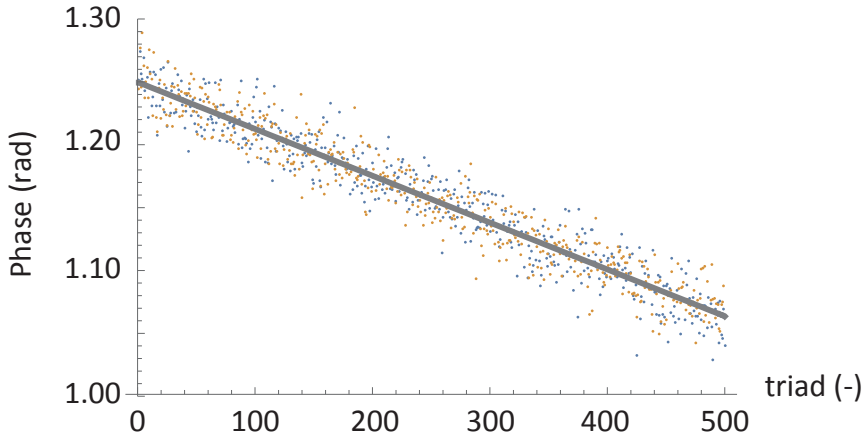


Figure 5.8. Estimated phase of the triad.

buffer (Figure 5.8).

When the smartphone was 2.0 m from the LED, the standard deviation of the regression was 0.015 rad, which is equivalent to 4.1 cm for an airborne acoustic wave. We use the standard error here to evaluate the precision of the estimated phase. Within 1 s (i.e., using at least 15 triads), the standard error of the regression becomes 1.1 cm for an airborne acoustic wave. With separations of 1.5 m and 1.0 m, the standard errors are 1.0 cm and 1.7 cm, respectively. The relationship between the separation and the precision, including analysis of the signal-to-noise ratio, should be investigated in detail in future work. Even though our previous study, which used a dedicated modulating waveform, showed a better standard error of 0.17 mm, the proposed system is sufficiently precise and is also applicable to VLC.

The timestamp of the image buffer and the time offset derived from the phase of the first line yield the reference time relative to the buffer's time. The precision of this reference time depends on the timing clock of the video camera. The standard deviation was 1.5 µs in this experiment.

Performance of Audio Signal Detection

We extracted the acoustic bursts from the video files and determined their positions on the audio track. The audio buffer size of the smartphone was 4,096 bytes and the timestamp of each buffer was recorded in a logging message. The emission time of each burst was calculated from the timestamp of the buffer to which the burst belonged and from the relative position of the burst in the buffer. The precision of the burst interval was 2 µs to 7 µs.

5.3.3. Summary

We have confirmed the feasibility of time synchronization using VLC. We successfully transmitted short messages and were able to obtain time references with practicable precision. The audio signals were also detected precisely. However, we did not align the video and audio tracks on the time axis in this experiment.

5.4. 3D ToA Localization Experiment using VLC Synchronization

In this section, we describe a 3D ToA localization experiment using VLC synchronization by a smartphone application. Alignment of video and audio frames is confirmed. The application works stand-alone.

5.4.1. Fast and Strict Inner Product Calculation for Phase Accordance Method

In the previous sections, we calculated the inner product of continuous time function $f(t)$ and $g(t)$ using discrete number series f_n and g_n as:

$$\langle f(t), g(t) \rangle = \frac{1}{T} \int_{-T/2}^{T/2} f(t)\overline{g(t)} \approx \frac{1}{N} \sum_{i=1}^{N} f_i \overline{g_i}. \tag{5.8}$$

Because this is the zeroth-order approximation of a continuous time function integral, we had to employ a high order oversampling to obtain accurate results.

As the sampling theorem shows we can convert a time series s_k into the equivalent continuous time function $s(t)$ like:

$$s(t) = \sum_k s_k \operatorname{sinc}(t - k). \tag{5.9}$$

Therefore, the inner product of $s(t)$ with a complex sinusoidal function $e^{j\omega t}$ as:

$$\langle s(t), e^{j\omega t} \rangle = \sum_k s_k \cdot \frac{1}{T} \int_{-T/2}^{T/2} \operatorname{sinc}(t - k) e^{-j\omega t} dt = \sum_k s_k \Gamma_k, \tag{5.10}$$

where

$$\Gamma_k = \frac{1}{T} \int_{-T/2}^{T/2} \operatorname{sinc}(t - k) e^{j\omega t} dt. \tag{5.11}$$

As Γ_k are constants, we can compute those integrals and store them in a table. Therefore, (5.10) can be carried out as a table-lookup manner.

5.4.2. Experimental Setup

We conducted an experiment to evaluate the proposed method. Figure 5.9 shows a diagram of the experimental system. In the transmitter (Figure 5.10), an LED light (PBT-170, JEFCOM) was placed to emit the modulated light. Although the LED light originally consisted of white LED elements and a driver circuit, the circuit was removed to modulate the current for the LED elements. The waveform signal of the LED modulation was generated by an arbitrary function generator (AFG3102, Tektronix). This signal drove a lab-built voltage-to-current converter that supplied electric current to the LED light.

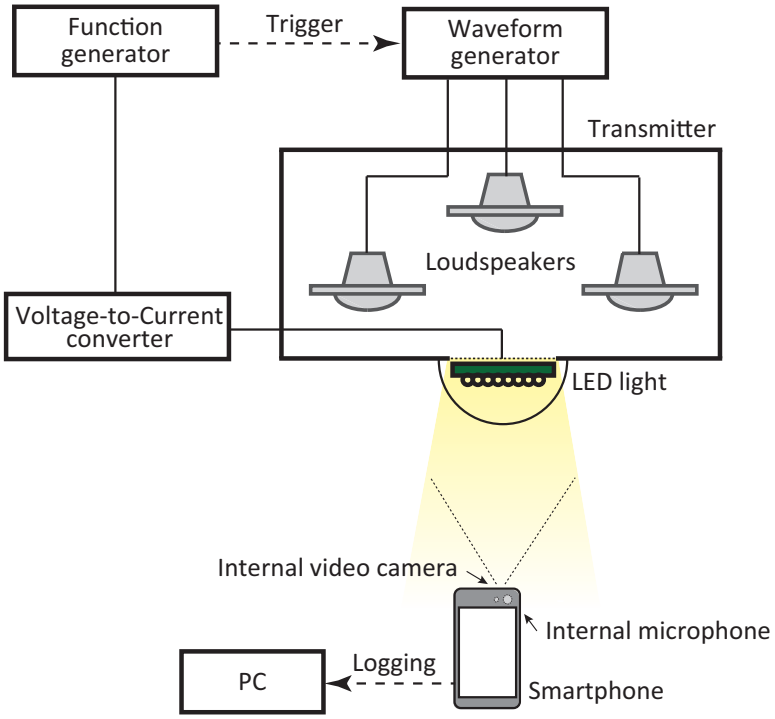


Figure 5.9. Experimental system for ToA localization using VLC synchronization.

Three loudspeakers were mounted on the transmitter for acoustic localization. The coordinates of the loudspeakers are shown in Table 5.1. The distance between any two of the three loudspeakers was 0.3 m.

Acoustic waves were emitted in 4 ms two-beat sync pattern through the loudspeakers at the beginning of each text character (Figure 5.11). Frequency ranges of each loudspeaker were 12750–13250 Hz, 13750–14250 Hz, and 14750–15250 Hz. FDM-PAM was used to determine the precise time of signal reception.

A smartphone (iPhone 6s Plus with iOS 10.2.1, Apple) was mounted on a tripod at a distance of 1.0 m or 1.5 m from the LED to capture video images. There were four localizing positions (Table 5.1). At each position, measurements were performed under two conditions — dark and bright. In the dark condition, the ceiling lights in the room were off and the Venetian blinds were

Figure 5.10. Transmitter for ToA localization uisng VLC synchronization.

Table 5.1. Coordinates of loudspeakers and microphone positions.

Device	Coordinates (x, y, z) (m)
Loudspeaker #1	(-0.150, 0.000, 1.165)
Loudspeaker #2	(0.000, 0.000, 0.905)
Loudspeaker #3	(0.300, 0.000, 0.905)
Microphone position #1	(0.000, -1.500, 0.905)
Microphone position #2	(-0.500, -1.500, 0.905)
Microphone position #3	(0.000, -1.000, 0.905)
Microphone position #4	(-0.500, -1.000, 0.905)

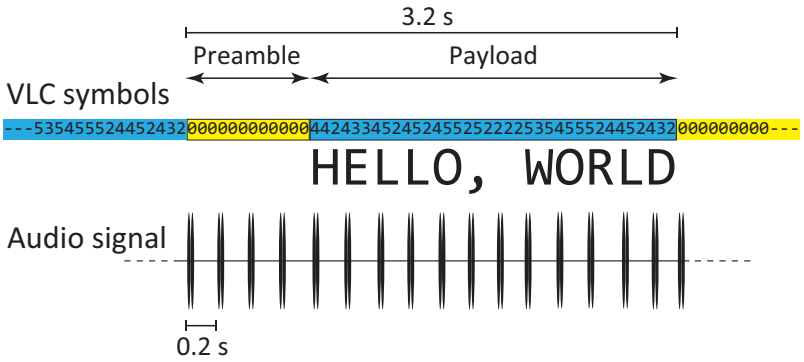


Figure 5.11. VLC message and audio signal.

closed, however the room was not completely dark. On the other hand, in the bright condition, the ceiling lights were on and the window blinds were not closed.

We developed a localization application to allow the smartphone to record the pixel values of two RoIs (RoI1 and RoI2). Because they were separated in the shutter-scanning direction, the detected phases of RoI1 and RoI2 were different. The frame rate of the video camera was 60 fps. Although demodulation was processed with the localization application, the recorded pixel values were sent as logging messages and analyzed offline by a personal computer (PC) to evaluate the performance statistically. Timestamps of the video frames were extracted from the sampling buffer of the smartphone through its application programming interface.

Audio signals captured by the smartphone's microphone were recorded alongside the video images as MPEG-4 files, which were transferred to the PC to calculate the stability of the received signals. The audio sampling rate was 48.0 ksps. Timestamps of the audio buffer were also extracted and recorded for analysis.

The LED's modulating signal was generated as follows (Figure 5.11). The signal consisted of a preamble and a payload. The preamble was used to detect the existence of a VLC message and to find the top frame of a triad. It comprised 12 "0" symbols of the 6-PSK. For the payload, we sent the text message "HELLO, WORLD". Each character in the text message was converted to a 6-bit binary code and divided into three 2-bit pieces that were assigned to the symbols "2"–"5". The symbol "1" was not used for data transmission but its existence was used as an integrity check on the receiver side. The payload of the transmission had 36 symbols, which means that the total length of the signal was 48 symbols. As each symbol occupied four video frames, the duration of the signal was 3.2 s. The signal was transmitted repeatedly, and the demodulation and the time synchronization process were performed after the entire VLC message was received. The video capture duration was 10 minutes for each measurement to calculate the statistics afterward.

5.4.3. Performance of VLC Demodulation and Time Synchronization

The pixel values captured by the smartphone application were demodulated as follows. We assume a sequence of pixel values $p_i (i = 1, 2, 3, \dots)$ in each RoI. We can make a series of triads as $t_1 = (p_1, p_2, p_3)$, $t_2 = (p_2, p_3, p_4)$, $t_3 = (p_3, p_4, p_5)$, $t_4 = (p_4, p_5, p_6)$, $\dots$. As mentioned before, there is a *right* triad that does not span two symbols in $t_1 \dots t_4$. A right triad appears every four triads. Triads in the preamble are used to select the correct sequence of triads in the payload.

A typical constellation diagram of the results is shown in Figure 5.12 (microphone position #1, dark condition). Here, the plotting plane is rotated so that the mean phase of RoI1's preamble becomes zero. We can see that the symbols are definitely separated, which means that the message was perfectly demodulated. The phase difference between RoI1 and RoI2 is due to the scanning speed of the rolling shutter. Using this speed and the phase of RoI1 or RoI2, we can estimate the phase of the first scanning line of the image buffer. Figure 5.13 shows the estimated phases of RoI1 and RoI2. Because the clock frequencies of the light modulation and the smartphone are different, the plots have slopes of around 40 ppm.

The standard deviations of the regression were 0.0049 rad for RoI1 and 0.048 rad for RoI2, which were equivalent to 13 mm for an airborne acoustic wave. We use the standard error here to evaluate the precision of the estimated phase. Within 1 s (i.e., using at least 15 triads), the standard error of the regression becomes 3.4 mm for an airborne acoustic wave. Even though

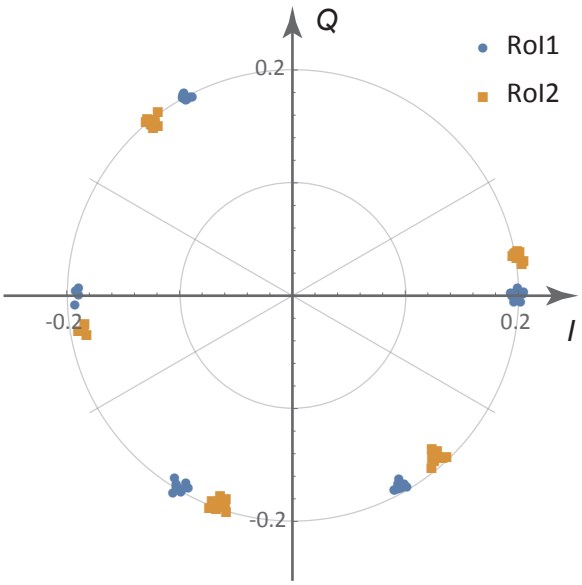


Figure 5.12. Constellation diagram of the results (microphone position #1, dark condition).

our previous study, which used a dedicated modulating waveform, showed a better standard error of 0.17 mm, the proposed system is sufficiently precise and is applicable to VLC.

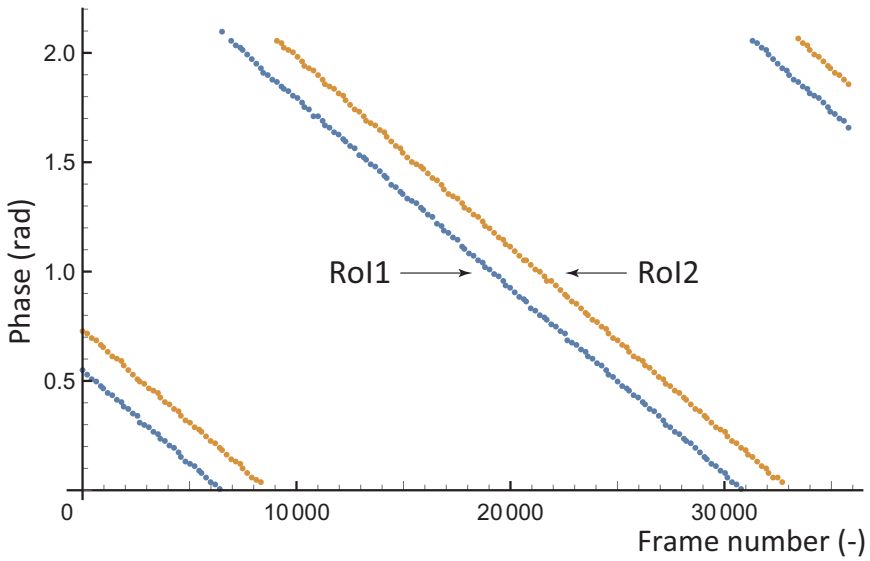


Figure 5.13. Estimated phase of the triad.

The timestamp of the image buffer and the time offset derived from the phase of the first line yield the reference time relative to the buffer's time. The precision of this reference time depends on the timing clock of the video camera. In this experiment, the standard deviation was 1.5 µs, which is equivalent to 0.5 mm for an airborne acoustic wave. This is much lower than the deviation of the RoI's phases.

5.4.4. Performance of Audio Signal Detection

We extracted the acoustic bursts from the audio buffer and determined their positions. The audio buffer size of the smartphone was 4,096 bytes and the timestamp of each buffer was recorded in a logging message. The emission time of each burst was calculated from the timestamp of the buffer to which the burst belonged and from the relative position of the burst in the buffer. The precision of the burst interval was 2 µs to 7 µs.

5.4.5. ToA Localization

Using the time reference extracted from the VLC message, the smartphone application measured the ToF of each sound wave. Then, it performed the 3D ToA localization. Analyzing the log messages of the localization application, the data were statistically summarized by the PC.

The results are shown in Table 5.2 and Figure 5.14. The overall standard deviations were around 100 mm under the dark condition and 100 mm to 200 mm under the bright condition. The offset errors were 100 mm to 150 mm for both cases. The standard deviation in the y-axis is larger than those of x- and z-axes. One reason of this error distribution is the frequency differences between the transmitter's clock and the smartphone's one. As mentioned in Subsection 5.4.3, the difference was 40 ppm; current localization application does not take account of it. As the acoustic wave arrived at the smartphone, the application soon calculated its ToF. On the other hand, the VLC symbols were demodulated when the whole message was received. The duration of the VLC message was 3.2 s; therefore, the time reference was 3.2 s old in the worst case, which would yield the offset equivalent to 0.044 m of an airborne acoustic wave. Using the time reference derived from the regression lines of Figure 5.13, the estimated positions have less variances shown in Figure 5.15 (standard deviations are 10–20 mm under the dark condition). We will introduce this correction in the next development.

Table 5.2. 3D ToA localization results.

Condition	Mean (m)	Standard deviation (m)		Offset
	(x, y, z)	(x, y, z)	Overall	(m)
#1, dark	(-0.017,-1.607, 0.875)	(0.016, 0.085, 0.010)	0.087	0.112
#2, dark	(-0.632,-1.558, 0.954)	(0.033, 0.102, 0.005)	0.107	0.153
#3, dark	(-0.103,-1.087, 0.872)	(0.006, 0.085, 0.010)	0.086	0.139
#4, dark	(-0.517,-1.070, 0.808)	(0.038, 0.109, 0.019)	0.117	0.121
#1, bright	(-0.023,-1.601, 0.873)	(0.018, 0.137, 0.013)	0.139	0.108
#2, bright	(-0.622,-1.535, 0.955)	(0.050, 0.164, 0.007)	0.172	0.136
#3, bright	(-0.104,-1.083, 0.873)	(0.007, 0.134, 0.015)	0.135	0.137
#4, bright	(-0.512,-1.054, 0.809)	(0.039, 0.115, 0.020)	0.123	0.111

5.5. Discussion

Here, we discuss the data length needed for our VLC localization system. Suppose that all transmitters have the same alignment and orientation of loudspeakers. We assume to send just

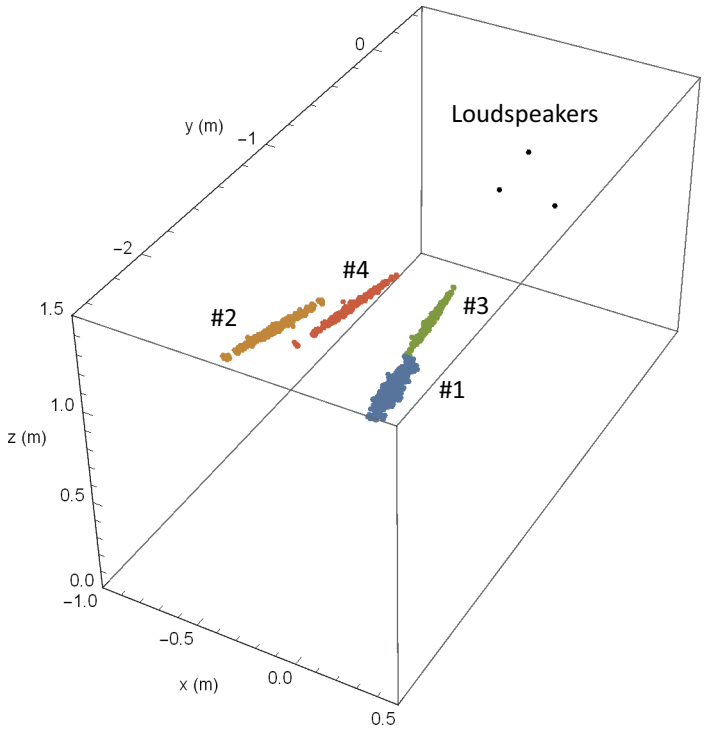


Figure 5.14. Localization results under dark condition.

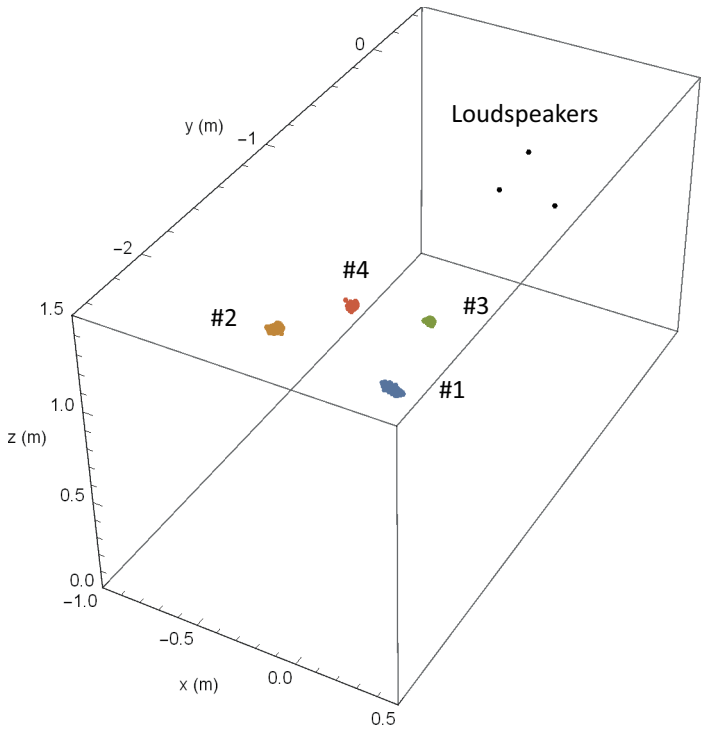


Figure 5.15. Estimated positions when completely synchronized under dark condition.

only the central coordinates of the transmitter. They can be notated using 4-bit 35 numeric characters with a precision of 3.1 mm according to the specification by Geospatial Information Authority of Japan (`http://ucopendb.gsi.go.jp/ucode/document/1_shiyo.pdf`). It takes 5.5 s to send them as the text message in the same setup as the experiment. The time is rather long, but can be shortened by using a shorter preamble and deploying other modulation schemes such as high-order quadrature amplitude modulation.

5.6. Summary and Future Work

We have confirmed the feasibility of 3D ToA localization using VLC. We successfully transmitted short messages and were able to obtain time references with practicable precision. The audio signals were also detected precisely. Finally, we conducted ToA-based 3D localization experiments and obtained the precision of 100 mm to 200 mm, which would be improved by compensating the phase estimation for time synchronization. We will cope with this issue in the future work.

There are other issues to be solved. Acoustic waves in this study have high frequencies and are almost inaudible for the human ear, however it might annoy young people. Furthermore, because the carrier frequency for VLC modulation is low, flicker is noticeable. We also deal with these issues.

6. Conclusion

6.1. Discussion and Future Work

There are some issues to be discussed on our methods.

First, we should deploy much higher frequencies for acoustic waves. We used 13 kHz to 17 kHz sound waves in our experiments according to our earlier experiments. However, these frequencies are rather audible especially for young people and children. To use higher frequencies, we should examine the frequency responses of the smartphone's microphones.

The frequency responses of the loudspeaker and the microphone, both magnitude and phase, are important factors for acoustic localization using FDM-PAM. If the magnitude is too low, we cannot acquire the signal. Moreover, if the phase delays are different between frequencies, they appear as time delays in PAM that utilizes phase information of the waves. The directivity of the frequency response also affects the accuracy of localization. We need calibration for practical use.

Next, flickering of the LED light modulation should be reduced. Although we have already developed a flicker-less VLC system, it is not applicable for time synchronization because there appear multiple time reference points in an image. Increasing the intensity of the LED light when the signal is "off", flickering is mitigated, however this is not a fundamental solution.

Below are other issues. Capturing video images consumes electric power. The system needs line-of-sight to the transmitter. Transmitter should be set at a higher position or on the ceiling. Time needed for synchronization should be reduced. Two regions-of-interest are needed for measuring the shutter-scanning speed. We should fix the smartphone's position during the synchronization process. Localization accuracy gets worse during walking. VLC messages should be assembled for localization.

6.2. Summary

We described some methods for smartphone localization. They use smartphone's internal sensors, a microphone for localization and a video camera for localization. Comparing RF signal sensors such as Wi-Fi and BLE, those sensors are scheduled to collect data regularly. Therefore, we can achieve good performance for localization.

Indoor localization is an emerging technique; hence evaluation methods have not been established yet. We conducted experiments to evaluate the performance at the distance of 2 m to 3 m in a small office room. We need to evaluate in larger spaces for practical use.

As for deploying our technique in real rooms and corridors, combination with PDR should be practicable. At least in Japan, there are emergency exit lights and loudspeakers with the separation of 10 m to 20 m. These devices might be used as base stations for time synchronization and precise acoustic localization. Even in the intermediate area between the stations where

localization signals are hard to capture, we can continuously keep localization process using PDR.

Here, we summarize our methods as follows.

First, we developed frequency division multiplexing phase accordance method (FDM-PAM). Using FDM-PAM, we can detect signals with a finer resolution than a sampling interval. Therefore, we will get precise ToF of sound waves for acoustic localization with high precision. We confirmed the feasibility of FDM-PAM through 2D TDoA localization experiments.

Next, we devised a time synchronization method using a modulated LED light and a video camera. Exploiting the rolling shutter effect of a CMOS image sensor, the resolution of the reference time detection is finer than the video image's row scanning interval, which is less than one thousandth of the frame rate. Our experimental results showed that within 1 s we obtained a time synchronization precision of 0.03 rows, which is equivalent to 0.17 mm of an airborne acoustic wave. With using this time synchronization technique, we conducted 3D ToA localization experiments. The results showed better performance than TDoA localization, which was calculated at the same measurements. Although the baselines of the transmitter beacons were short as 0.3 m, the precision of the ToA localization was not more than 10 mm at the distance of 2.0 ms.

Finally, we proposed a time synchronization technique using VLC. VLC is inherently realized on a synchronized channel; demodulation of VLC yields time synchronization. In addition, VLC conveys messages; parameters needed for localization, such as coordinates of the sender beacons, can be transmitted through VLC. We demonstrated a 3D ToA localization application using VLC, which had sufficient precision for practical applications.

A. Theoretical Analysis of Modulating Light and Video Images

A.1. Fourier Analysis of Light Modulation

Suppose that the light source $b(t)$ is modulated periodically at the frequency 1. We assume that we have a video camera with a frame rate m and a duty cycle η for the charge integration of the image sensor. The video camera's shutter works as a rectangular window of width η/m:

$$w(t) = \begin{cases} 1 & \left(-\frac{\eta}{2m} \le t \le \frac{\eta}{2m}\right) \\ 0 & (\text{otherwise}) \end{cases}. \tag{A.1}$$

We consider m consecutive frames as one image $B(x)$, where x expresses continuous values on the interval $[0, 1)$. The pixel value at time $t = x$ is

$$B(x) = \int_x^{x+1} w\left(t - x - \frac{\eta}{2m}\right) b(t)\, dt. \tag{A.2}$$

Considering $w(-t) = w(t)$ and $B(x + 1) = B(x)$,

$$B(x) = \int_0^1 w\left(x + \frac{\eta}{2m} - t\right) b(t)\, dt \tag{A.3}$$

holds. The right-hand side of this equation is the convolution of $w(x + \eta/2m)$ and $b(x)$. Now, we describe $B(x)$ using complex Fourier coefficients $\tilde{B}_k$:

$$B(x) = \sum_{k=-\infty}^{\infty} \tilde{B}_k e^{j2\pi kx}. \tag{A.4}$$

Equation (A.3) shows that $\tilde{B}_k$ are the product of complex Fourier coefficients of $w(x + \eta/2m)$ and $b(x)$.

We define $\tilde{w}_k$ and $\tilde{b}_k$ as complex Fourier coefficients of $w(x)$ and $b(x)$, respectively. For rectangular window $w(x)$, its complex Fourier coefficients are

$$\tilde{w}_k = \frac{\eta}{m} \operatorname{sinc} \frac{\pi k \eta}{m}. \tag{A.5}$$

Time-shifting $w(x)$ by $-\eta/2m$, we can get the complex Fourier coefficients of $w(x + \eta/2m)$ as

$$e^{-j2\pi k\left(-\frac{\eta}{2m}\right)} \cdot \frac{\eta}{m} \operatorname{sinc} \pi k \frac{\eta}{m} = \frac{\eta}{m} e^{j\pi k \frac{\eta}{m}} \operatorname{sinc} \pi k \frac{\eta}{m}. \tag{A.6}$$

Finally, we get

$$\tilde{B}_k = \frac{\eta}{m} e^{j\pi k \frac{\eta}{m}} \operatorname{sinc} \pi k \frac{\eta}{m} \tilde{b}_k. \tag{A.7}$$

A.2. Experimental Setup and Theoretical Image Pattern

The complex Fourier coefficients $\tilde{b}_k$ of the square wave $b(x)$ are

$$\tilde{b}_k = \frac{1}{2} e^{-j\frac{\pi k}{2}} \operatorname{sinc} \frac{\pi k}{2}, \tag{A.8}$$

which is derived in a similar way to getting the complex Fourier coefficients of $w(x + \eta/2m)$. In the experiment, we analyzed two consecutive frames and the exposure time was the same as the frame interval. Therefore, $m = 2$ and $\eta = 1$. Substituting these values and equation (A.8) into equation (A.7), we obtain

$$\tilde{B}_k = \frac{1}{4} \operatorname{sinc}^2 \frac{\pi k}{2}. \tag{A.9}$$

This means that $B(x)$ is a triangle wave.

B. PDOP for ToA and TDoA

We can derive the PDOP for ToA and TDoA as follows [29, 25]. Firstly, we describe the error covariance matrix of the ToA measurement. Let

$$R_i = ct_i + \varepsilon_i \tag{B.1}$$

be the measured distance between the ith loudspeaker's position $\boldsymbol{s_i} = (x_i, y_i, z_i)^{\mathrm{T}}$ and the smartphone's position $\boldsymbol{x} = (x, y, z)^{\mathrm{T}}$, where t_i is the measured time of flight from the ith loudspeaker, c is the speed of sound, and ε_i is an error in this measurement. Distance R_i is described as

$$\begin{aligned} R_i &= \|\boldsymbol{s}_i - \boldsymbol{x}\| \\ &= \sqrt{(x_i - x)^2 + (y_i - y)^2 + (z_i - z)^2} + \varepsilon_i. \end{aligned} \tag{B.2}$$

Here, we assume that we have an initial value of $\boldsymbol{x}$ as $\boldsymbol{x}_0 = (x_0, y_0, z_0)^{\mathrm{T}}$. Then, the first-order approximation of R_i around $\boldsymbol{x} = \boldsymbol{x}_0$ is

$$R_i = \|\boldsymbol{s}_i - \boldsymbol{x}_0\| + (\Delta\boldsymbol{x})^{\mathrm{T}} \left.\frac{\partial R_i}{\partial \boldsymbol{x}}\right|_{\boldsymbol{x}=\boldsymbol{x}_0} + \varepsilon_i, \tag{B.3}$$

where

$$\Delta\boldsymbol{x} = \boldsymbol{x} - \boldsymbol{x}_0 = (\Delta x, \Delta y, \Delta z)^{\mathrm{T}}. \tag{B.4}$$

If we now introduce R_{0i} as

$$\begin{aligned} R_{0i} &= \|\boldsymbol{s}_i - \boldsymbol{x}_0\| \\ &= \sqrt{(x_i - x_0)^2 + (y_i - y_0)^2 + (z_i - z_0)^2}, \end{aligned} \tag{B.5}$$

then

$$\begin{aligned} \Delta R_i =& R_i - R_{0i} \\ =& \frac{-(x_i - x_0)}{R_{0i}}\Delta x + \frac{-(y_i - y_0)}{R_{0i}}\Delta y + \frac{-(z_i - z_0)}{R_{0i}}\Delta z + \varepsilon_i. \end{aligned} \tag{B.6}$$

Using n loudspeakers, we can make a matrix equation

$$\begin{pmatrix} \Delta R_1 \\ \vdots \\ \Delta R_n \end{pmatrix} = \begin{pmatrix} \dfrac{-(x_1 - x_0)}{R_{01}} & \dfrac{-(y_1 - y_0)}{R_{01}} & \dfrac{-(z_1 - z_0)}{R_{01}} \\ \vdots & \vdots & \vdots \\ \dfrac{-(x_n - x_0)}{R_{0n}} & \dfrac{-(y_n - y_0)}{R_{0n}} & \dfrac{-(z_n - z_0)}{R_{0n}} \end{pmatrix} \begin{pmatrix} \Delta x \\ \Delta y \\ \Delta z \end{pmatrix} + \begin{pmatrix} \varepsilon_1 \\ \vdots \\ \varepsilon_n \end{pmatrix}, \tag{B.7}$$

which is of the form

$$\boldsymbol{y} = A\boldsymbol{x} + \boldsymbol{\varepsilon}, \tag{B.8}$$

where $\boldsymbol{y}$, $\boldsymbol{x}$, and $\boldsymbol{\varepsilon}$ are $n \times 1$ vectors and A is an $n \times n$ matrix. Note that $\boldsymbol{x}$ has been redefined.

We assume that the components of the error vector $\boldsymbol{\varepsilon}$ are identically distributed and independent, and that their variances are equal to σ_0. In this case, the covariance matrix of $\boldsymbol{\varepsilon}$ is

$$\Sigma = \mathrm{E}[\boldsymbol{\varepsilon}\boldsymbol{\varepsilon}^{\mathrm{T}}] = {\sigma_0}^2 I, \tag{B.9}$$

where $\mathrm{E}[\cdot]$ expresses the expected value and I is an identity matrix. The weighted least-squares estimation of $\boldsymbol{x}$ is

$$\hat{\boldsymbol{x}} = \left(A^{\mathrm{T}}\Sigma^{-1}A\right)^{-1} A^{\mathrm{T}}\Sigma^{-1}\boldsymbol{y} \tag{B.10}$$

and its covariance matrix is

$$V = \left(A^{\mathrm{T}}\Sigma^{-1}A\right)^{-1}. \tag{B.11}$$

In the TDoA measurement, we can only get the time difference $R_i - R_j$ or $\Delta R_i - \Delta R_j$. If we choose the first loudspeaker's signal as the reference, then the matrix equation is

$$\boldsymbol{y}_d = A_d\boldsymbol{x} + \boldsymbol{\varepsilon}_d, \tag{B.12}$$

where

$$\boldsymbol{y}_d = M\boldsymbol{y}, \tag{B.13}$$
$$A_d = MA, \tag{B.14}$$
$$\boldsymbol{\varepsilon}_d = M\boldsymbol{\varepsilon}, \tag{B.15}$$
$$M = \begin{pmatrix} -1 & 1 & & 0 \\ \vdots & & \ddots & \\ -1 & 0 & & 1 \end{pmatrix}. \tag{B.16}$$

The weighted least-squares estimation of $\boldsymbol{x}$ is

$$\hat{\boldsymbol{x}}_d = \left({A_d}^{\mathrm{T}}{\Sigma_d}^{-1}A_d\right)^{-1} {A_d}^{\mathrm{T}}{\Sigma_d}^{-1}\boldsymbol{y}_d \tag{B.17}$$

and its covariance matrix is

$$V_d = \left({A_d}^{\mathrm{T}}{\Sigma_d}^{-1}A_d\right)^{-1}, \tag{B.18}$$

where

$$\Sigma_d = \mathrm{E}\left[(M\boldsymbol{\varepsilon})(M\boldsymbol{\varepsilon})^{\mathrm{T}}\right] = {\sigma_0}^2 \begin{pmatrix} 2 & 1 & \dots & 1 \\ 1 & 2 & \ddots & \vdots \\ \vdots & \ddots & \ddots & 1 \\ 1 & \dots & 1 & 2 \end{pmatrix}. \tag{B.19}$$

Matrices V and V_d are of the form

$$\begin{pmatrix} {\sigma_x}^2 & \sigma_{xy} & \sigma_{xz} \\ \sigma_{yx} & {\sigma_y}^2 & \sigma_{yz} \\ \sigma_{zx} & \sigma_{zy} & {\sigma_z}^2 \end{pmatrix}. \tag{B.20}$$

Here, σ_x/σ_0, σ_y/σ_0, and σ_z/σ_0 express the magnification of the measurement error in the x-, y-, and z-directions, respectively. Now we can drive the PDOP as

$$\text{PDOP} = \frac{\sqrt{\sigma_x^2 + \sigma_y^2 + \sigma_z^2}}{\sigma_0}. \tag{B.21}$$

C. Lab-made Devices and Programs

C.1. Audio Amplifier

A 4-channel audio amplifier (Figure C.1) was used to drive loudspeakers. It was made of four audio amplifier kits ("Toshiba TA7252AP Audio Amplifier Kit (Monaural) 5.9W", Akizuki Denshi Tsusho). The circuit diagram is shown in Figure C.2.

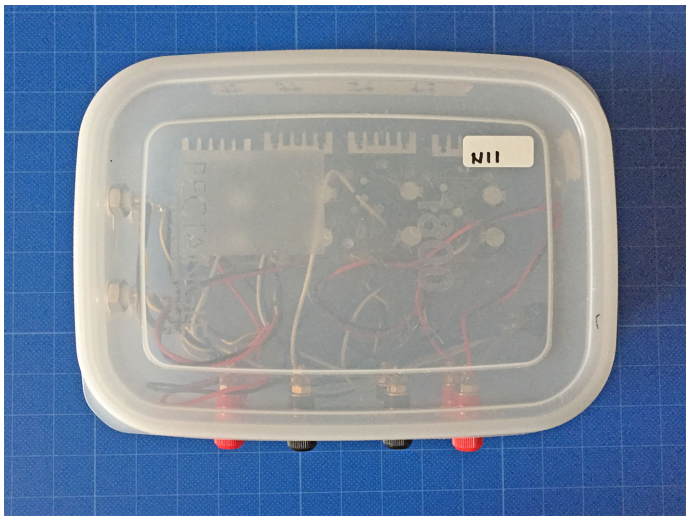


Figure C.1. 4-channel amplifier.

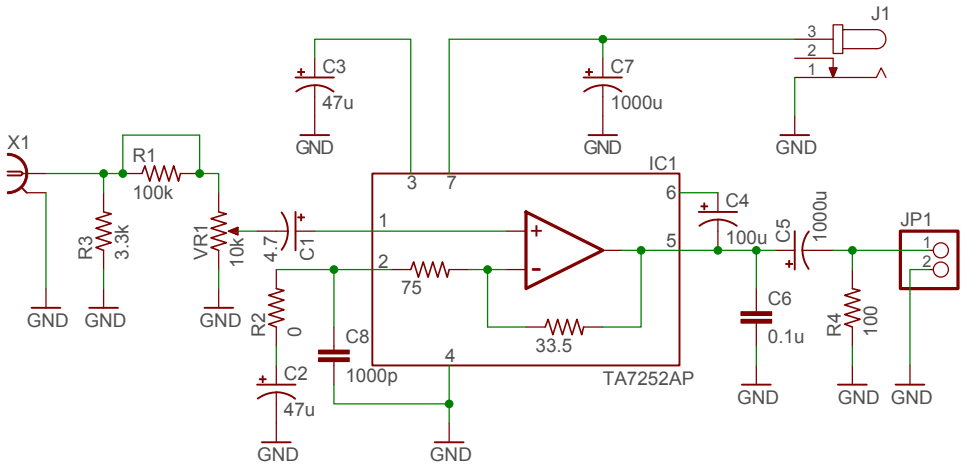


Figure C.2. Circuit diagram of the 4-channel amplifier.

C.2. Audio Waveform Generator

A programmable audio waveform generator (Figure C.3) was used. It is controlled by a microcontroller board ("ATmega168/328 *micom* board kit", Akizuki Denshi Tsusho) that is compatible with Arduino UNO. Its circuit diagram is shown as Figure C.4.

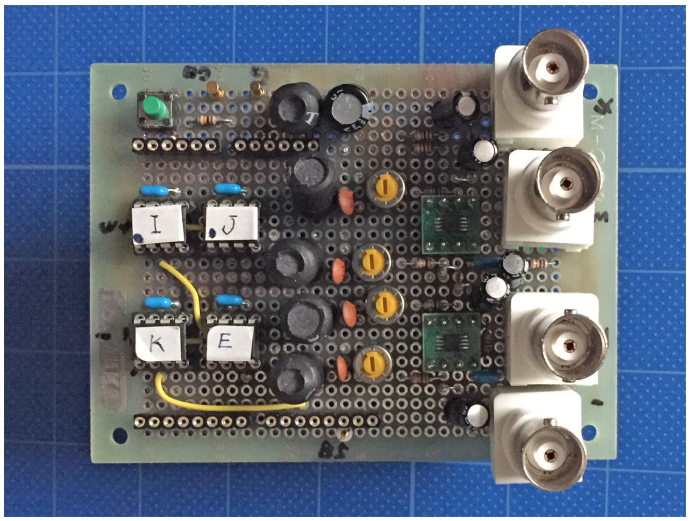


Figure C.3. Audio waveform generator.

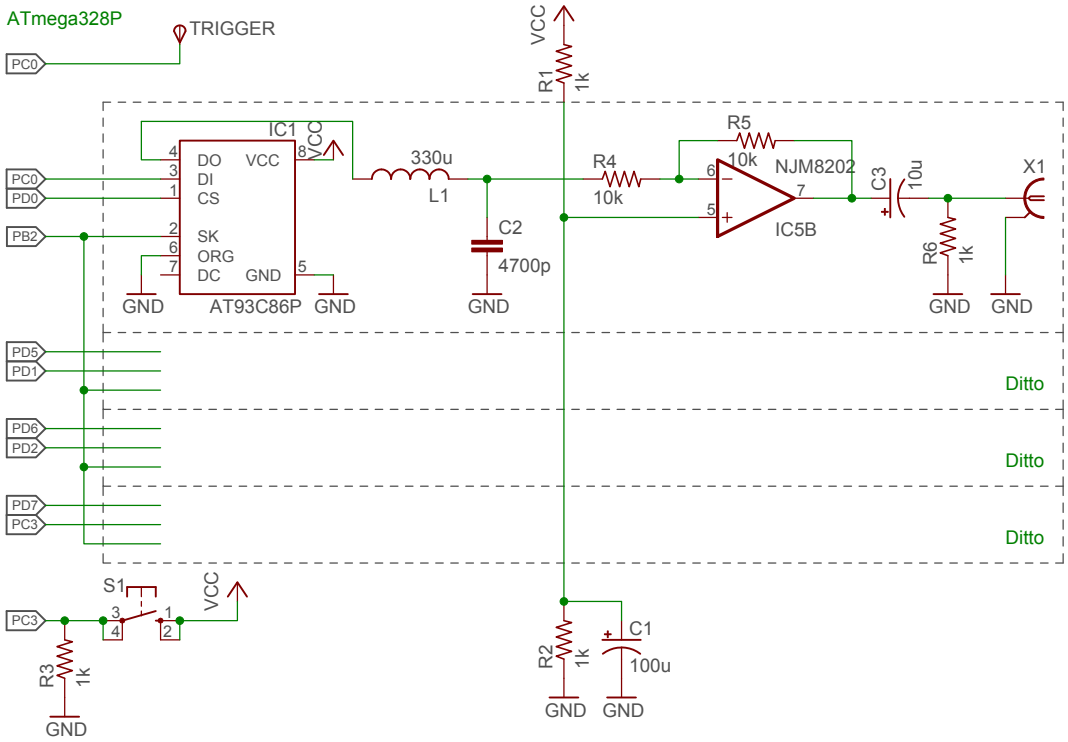


Figure C.4. Circuit diagram of the audio waveform generator.

C.3. Program for Generating ROM Data for Audio Signal Generator

To generate ROM data for the audio signal generator, the script below was used. It works on Mathematica 11.

Making ROM data for audio waveform generator

Number of bits for the entire data

```
romsize = 16 384;
```

Length of the sync pattern (msec)

```
dur = 4 / 1000;
```

Clock frequency for ROM

```
clock = 2*^6;
```

Number of bits for the sync pattern

```
num = clock dur
```

```
8000
```

Number of bits before the sync pattern

```
prelen = 2 / 1000 clock
```

```
4000
```

Number of bits after the sync pattern

```
postlen = romsize - prelen - num
```

```
4384
```

Frequencies for the sync pattern

```
{f1, f2} = {13 750, 14 250};
```

```
f2 - f1
```

```
500
```

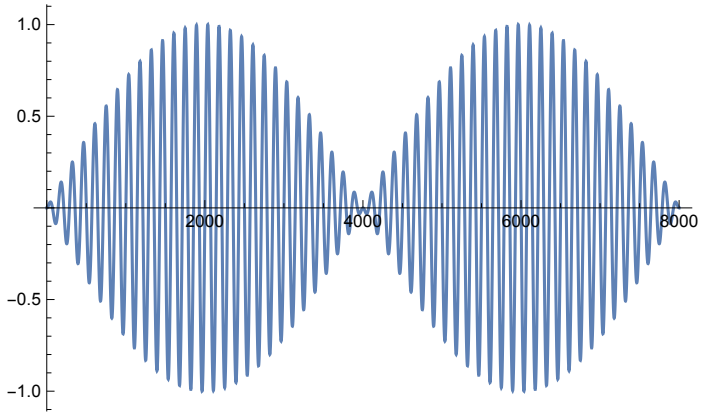
Sync pattern

```
makesync[fa_, fb_, duration_, n_, orig_] :=
 1 / 2 Table[Sin[2 Pi fa duration k / n] + Sin[2 Pi fb duration k / n], {k, orig, orig + n - 1}]
```

```
syncf1f2 = makesync[f1, f2, dur, num, -num / 4];
```

Figure C.5. ROM data generation program. (1/3)

```
ListLinePlot[syncf1f2, Joined → True]
```



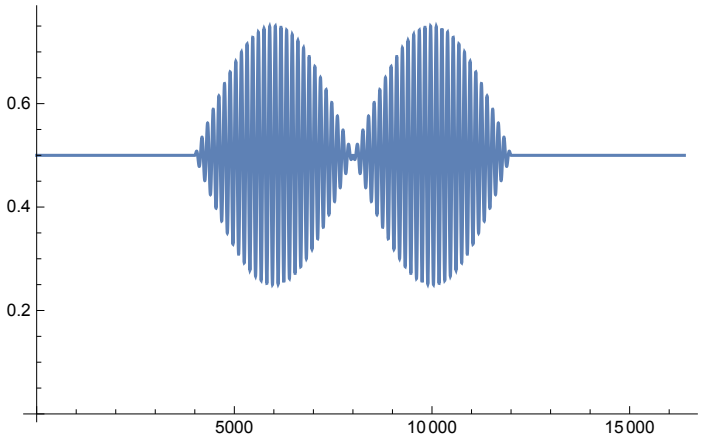
Delta sigma Modulation

```
deltasigmaLmap[sig_] := Module[{int1 = 0.5, last = 0},
  (int1 += # - last;
     last = If[int1 > 0.5, 1, 0]) & /@ sig]
```

Waveform

```
sigf1f2 = 1 / 2 Join[Table[1, {prelen}], (1 + 5 / 10 syncf1f2), Table[1, {postlen}]];

ListPlot[sigf1f2, Joined → True, PlotRange → All]
```



Binary data

```
dsoutf1f2 = deltasigmaLmap[sigf1f2];

Length[dsoutf1f2]
```

16 384

```
dsoutf1f2[[ ;; 10]]
```

{1, 0, 1, 0, 1, 0, 1, 0, 1, 0}

Figure C.6. ROM data generation program. (2/3)

```
dsoutf1f2[[-10 ;;]]
{1, 0, 1, 0, 1, 0, 1, 0, 1, 0}
```

Byte-Packed data

```
ivect = Table[2^k, {k, 7, 0, -1}]
{128, 64, 32, 16, 8, 4, 2, 1}

octf1f2 = Partition[dsoutf1f2, 8];

v1f1f2 = #.ivect & /@ octf1f2;

Length[v1f1f2]
2048
```

Verification

```
octf1f2[[ ;; 5]]
{{1, 0, 1, 0, 1, 0, 1, 0}, {1, 0, 1, 0, 1, 0, 1, 0},
 {1, 0, 1, 0, 1, 0, 1, 0}, {1, 0, 1, 0, 1, 0, 1, 0}, {1, 0, 1, 0, 1, 0, 1, 0}}

v1f1f2[[ ;; 5]]
{170, 170, 170, 170, 170}
```

File output

```
Export["rom-sample.dat", v1f1f2, "UnsignedInteger8"];
```

For verifying on ROM writer

Hexadecimal notation

```
BaseForm[#, 16] & /@ v1f1f2[[ ;; 16]]
{aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16, aa16}
```

Checksum

```
BaseForm[Mod[Total[#], 2^16], 16] & /@ {v1f1f2}
{aa7b16}
```

Figure C.7. ROM data generation program. (3/3)

C.4. Program for Audio Waveform Generator

To control the audio signal generator, an assembler program C.1 was used. The clock input for each ROM is connected to the PB2 output of the ATmega328P microcontroller. The output of PB2 alternates 0 and 1 every 10 clock cycles. The clock frequency is 20 MHz; therefore, each ROM is driven by 2 MHz clock.

Listing C.1 Program for audio waveform generator.

```
/*
 * audio_sig_gen.asm
 *
 * 2015-04-03
 */

/*
 * sysgena.asm
 *
 *  Created: 2013/07/18 9:08:42
 *   Author: akiyama
 */

.equ    PB5NUM = 32
.equ    PB5ON = 32
.equ    PB5OFF = 0

.equ    PB2NUM = 4
.equ    PB2ON = 4
.equ    PB2OFF = 0

.equ    CH1CS = 1
.equ    CH2CS = 2
.equ    CH3CS = 4
.equ    CH4CS = 8
.equ    CHACS = CH1CS | CH2CS | CH3CS | CH4CS

.equ    CH1DI = 16
.equ    CH2DI = 32
.equ    CH3DI = 64
.equ    CH4DI = 128
.equ    CHADI = CH1DI | CH2DI | CH3DI | CH4DI

;; ROMs are organized as 2048 x8 (11-bit address)

.equ    CKLEN = 1 + 11 + (2000 + 9071 + 2000)   ; 2MHz * 4ms = 8000b

.def    W = R16
.def    CKON = R17
.def    CKOFF = R18
```

```
.def    CCKH = r25
.def    CCKL = r24

.org    0x0000
        rjmp    RESET

.org    PCI1addr
        rjmp    OUT_AUDIO

RESET:

        cli

; set Stack Pointer

        ldi     W,      high(RAMEND)
        out     SPH,    W
        ldi     W,      low(RAMEND)
        out     SPL,    W

; set PB2 and PB5 as output pins

        ldi     W,      PB2NUM | PB5NUM
        out     DDRB,   W

        ldi     W,      0
        out PORTB,      W

; set PD0-PD7 as output pins

        ldi     W,      255
        out     DDRD,   W

; preset clock on/off values for PB2 and PB5

        ldi     CKON,   PB2ON | PB5ON
        ldi     CKOFF,  PB2OFF | PB5OFF

; set PC0-PC6 pull-uped

        ldi     W,      0
        out     DDRC,   W

        ldi     W,      127
        out     PORTC,  W

;       sbi     PORTC,  1 << PINC0
```

```
;
; Pin Change Interrupt
;

; enable PCINT8(PC0)

        lds     W,      PCMSK1
        ori     W,      1 << PCINT8
        sts     PCMSK1, W

; enable Pin Change Interrupt 1 (PCINT[14:8])

        lds     W,      PCICR
        ori     W,      1 << PCIE1
        sts     PCICR,  W

; enable global interrupt

        sei

; wait for interrupt

INF_LOOP:
        nop
        rjmp    INF_LOOP

;
; interrupt routine for emiting pulse
;

OUT_AUDIO:

; disable global interrupt

        cli

; if the pin is LOW, then return,
; because the interrupt was at the falling edge.

        sbis    PINC,   PCINT8
        reti

; output '1'

        ldi     W,      CHACS | CHADI
        out     PORTD,  W
        nop
```

```
        nop

        out     PORTB,  CKON            ; 1 (#Clocks)
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1

        out     PORTB,  CKOFF           ; 1
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1

; output '1'

        out     PORTB,  CKON            ; 1
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1
        nop                             ; 1

        out     PORTB,  CKOFF           ; 1
        nop                             ; 1
        nop                             ; 1

; set output data as '0'

        ldi     W,      CHACS           ; 1
        out     PORTD,  W               ; 1

; output '0's (ROM address) and read out ROM

        out     PORTB,  CKON            ; 1
        ldi     CCKH,   high(CKLEN)     ; 1
        ldi     CCKL,   low(CKLEN)      ; 1
        nop                             ; 1
        nop                             ; 1

        out     PORTB,  CKOFF           ; 1
        sbiw CCKH:CCKL, 1               ; 2
        nop                             ; 1
        nop                             ; 1

Loop:
        out     PORTB,  CKON            ; 1
        nop                             ; 1
        nop                             ; 1
```

```
        nop                                 ; 1
        nop                                 ; 1

        out     PORTB,  CKOFF               ; 1
        sbiw CCKH:CCKL, 1                   ; 2
        brbc    1,      Loop                ; 1/2

; set ROMs' CS off

        ldi     W,      0
        out     PORTD,  W       ; CS -> 0

        reti
```

C.5. LED Floodlight

An LED floodlight (Figure C.8) was used for time synchronization. LEDs were mounted on a circuit board (AE-LED56V2, a universal board for 8 series and 7 parallel LEDs, Akizuki Denshi Tsusho), which mas modified to 4 series and 14 parallel (Figure C.9).

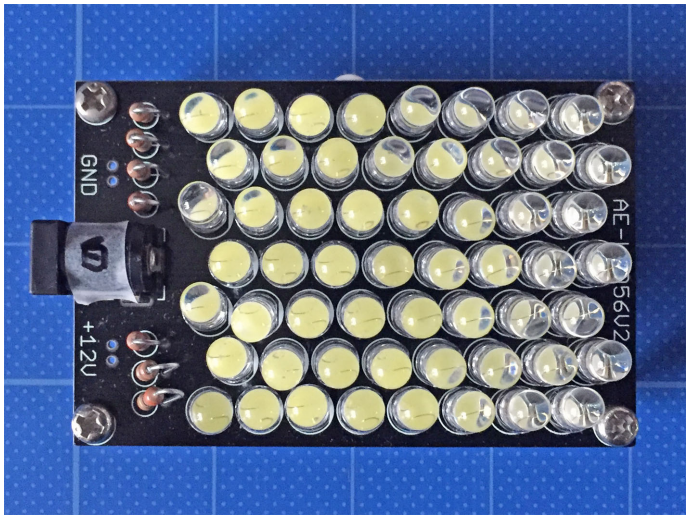


Figure C.8. LED floodlight.

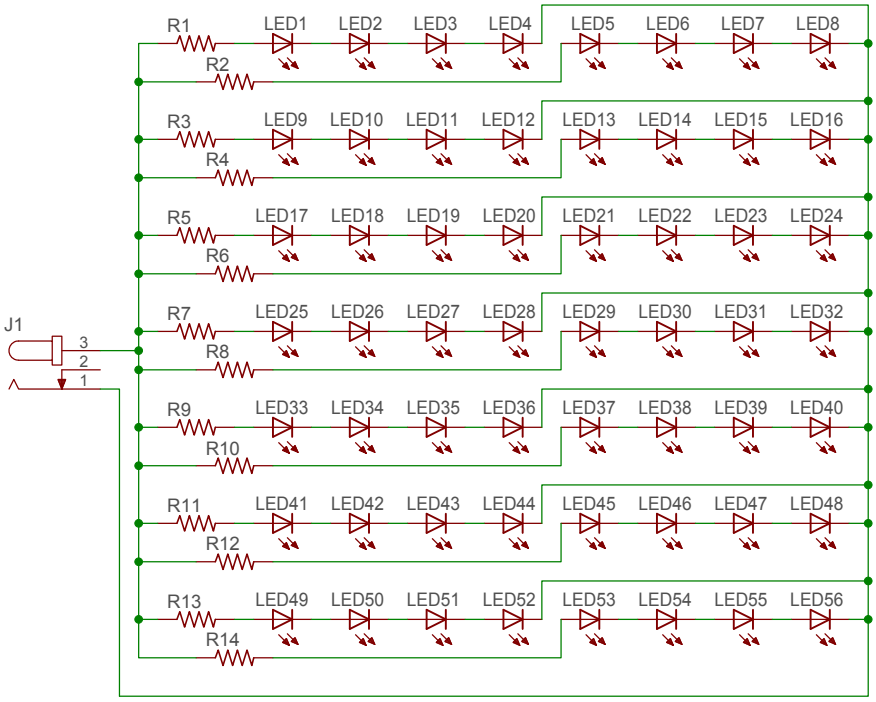


Figure C.9. Circuit diagram of the LED floodlight.

The floodlight was driven by a driver shown in Figure C.10. Figure C.11 shows the circuit diagram of the LED floodlight driver.

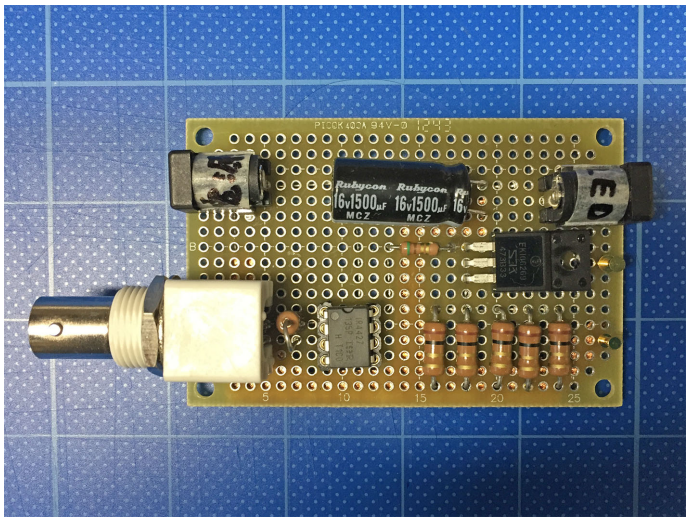


Figure C.10. LED floodlight driver.

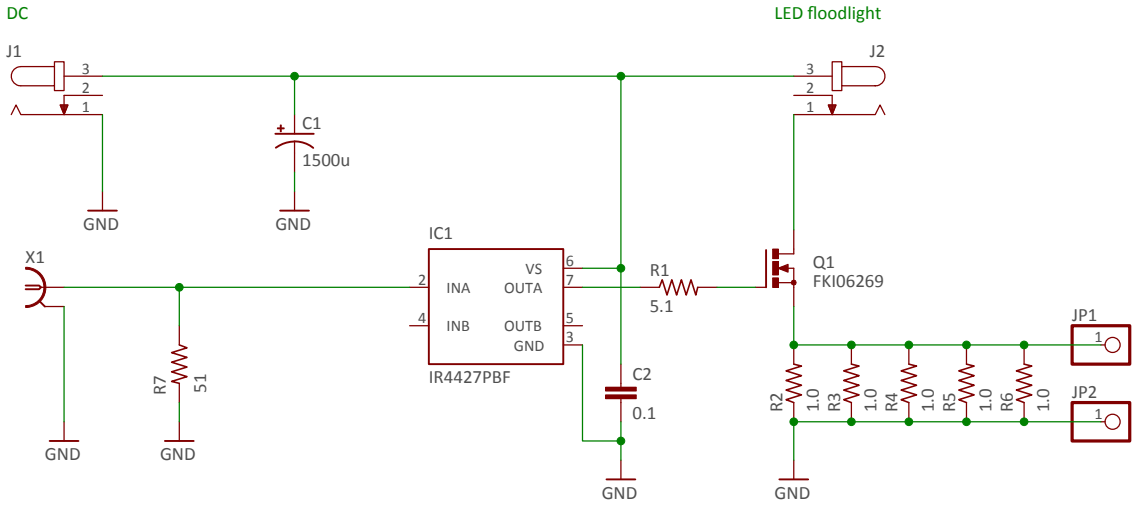


Figure C.11. Circuit diagram of the LED floodlight driver.

C.6. Voltage-to-current Converter for LED Light

A voltage-to-current converter (Figure C.12) was used to drive the LED light.

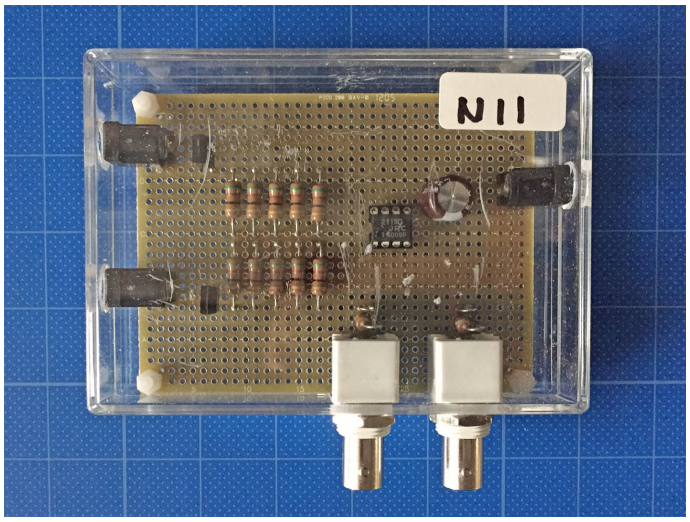


Figure C.12. Voltage-to-current converter for LED light.

Figure C.13 is the circuit diagram of the voltage-to-current converter.

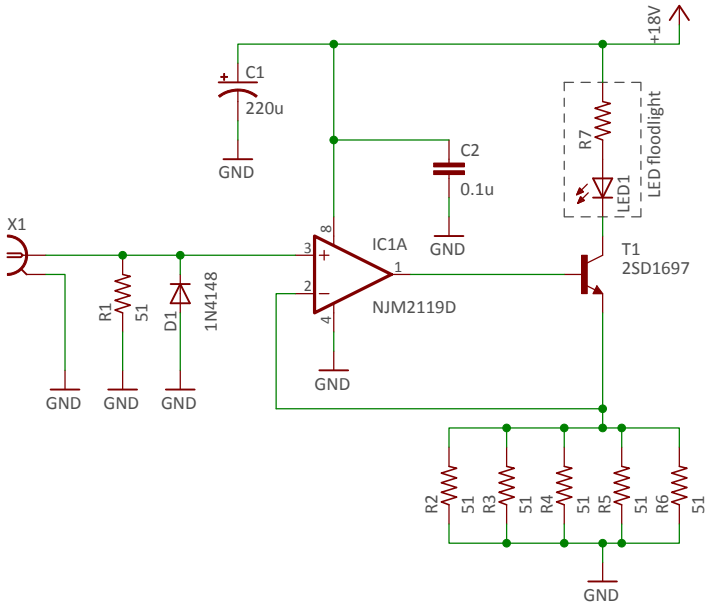


Figure C.13. Circuit diagram of the voltage-to-current converter.

Acknowledgements

The author really thanks the advisors and the committee: Professor HASHIZUME Hiromichi, Associate Professor GOTODA Hironobu, Professor TAKASU Atsuhiro, Professor JI Yusheng, Associate Professor ONO Nobutaka, Associate Professor KODAMA Kazuya, and Professor SUGIMOTO Masanori at Hokkaido University.

My advisor Professor Hashizume has arranged the schedule and conducted me to obtain the Ph.D. degree. He sometimes showed me theoretical insights into phenomena although I tend to prefer intuitive explanations. He also gave me many suggestions for experimental methods and devices.

One of my sub-advisor Associate Professor Gotoda asked me simple questions that let me reach another angle of view. His words let me notice important points.

The other sub-advisor Professor Takasu gave me strict questions, which trained me very hard. He showed me view points from other fields.

Professor Ji gently and strictly watched me. Her words suggested a way of extending my study and applying newly-developed technologies.

Associate Professor Ono taught me many technical methods especially on acoustic signal processing. His suggestions refine my methodologies and theories.

Associate Professor Kodama gave me some comments, which are simple but point out the cores of the problems. His words remind me the essential part of the study.

Professor Sugimoto held meetings to discuss issues of my study. He showed me ways to make presentations in conferences and to write papers.

I also thank alumni, graduate students, and undergraduate students at Hokkaido University who studied with me and discussed at meetings. Those of them include my co-authors of papers Mr. NAKAMURA Masanari, Mr. KUMAKI Hayato, Mr. SUWABE Kai, Mr. SAITO Taishi, Mr. SHIMADA Shota, Mr. KUDO Koki, and Mr. AMAYA Shogo.

Bibliography

[1] Omar Ait-Aider, Nicolas Andreff, Jean Marc Lavest, and Philippe Martinet. Exploiting Rolling Shutter Distortions for Simultaneous Object Pose and Velocity Computation Using a Single View. In *Fourth IEEE International Conference on Computer Vision Systems (ICVS'06)*, pages 35–35, New York, NY, USA, Jan 2006.

[2] Paramvir Bahl and Venkata N. Padmanabhan. 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 (Cat. No.00CH37064)*, volume 2, pages 775–784 vol.2, Tel Aviv, Israel, Mar 2000.

[3] Dennis A. Bohn. Environmental Effects on the Speed of Sound. *J. Audio Eng. Soc*, 36(4):223–231, Apr 1988.

[4] Wing Kwong Chung, Ning Cui, Xu Chen, Tin Lun Lam, Yangsheng Xu, and Yongsheng Ou. Optimized methodologies for augmented reality markers based localization. In *2014 IEEE International Conference on Robotics and Biomimetics (ROBIO 2014)*, pages 1811–1816, Bali, Indonesia, Dec 2014.

[5] Bert Cox, Geoffrey Ottoy, and Lieven De Strycker. Development of an Optoacoustic Distance Measurement System. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcala de Henares, Spain, Oct 2016.

[6] Paolo Dabove. What are the actual performances of GNSS positioning using smartphone technology? *Inside GNSS*, pages 35–37, November/December 2014.

[7] Christos Danakis, Mostafa Afgani, Gordon Povey, Ian Underwood, and Harald Haas. Using a CMOS camera sensor for visible light communication. In *2012 IEEE Globecom Workshops*, pages 1244–1248, Anaheim, CA, USA, Dec 2012.

[8] John M. Dow, Ruth E. Neilan, and Chris Rizos. The International GNSS Service in a changing landscape of Global Navigation Satellite Systems. *Journal of Geodesy*, 83(3):191–198, Mar 2009.

[9] Paolo Ferrari, Alessandra Flammini, Stefano Rinaldi, Andrea Bondavalli, and Francesco Brancati. Experimental Characterization of Uncertainty Sources in a Software-Only Synchronization System. *IEEE Transactions on Instrumentation and Measurement*, 61(5):1512–1521, May 2012.

[10] Viacheslav Filonenko, Charlie Cullen, and James Carswell. Investigating ultrasonic positioning on mobile phones. In *2010 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–8, Zurich, Switzerland, Sept 2010.

[11] Davide Della Giustina, Paolo Ferrari, Alessandra Flammini, and Stefano Rinaldi. Experimental characterization of time synchronization over a heterogeneous network for Smart Grids. In *2013 IEEE International Workshop on Applied Measurements for Power Systems (AMPS)*, pages 132–137, Aachen, Germany, Sept 2013.

[12] Carles Gomez, Joaquim Oller, and Josep Paradells. Overview and Evaluation of Bluetooth Low Energy: An Emerging Low-Power Wireless Technology. *Sensors*, 12(9):11734–11753, Aug 2012.

[13] Robert Harle. A Survey of Indoor Inertial Positioning Systems for Pedestrians. *IEEE Communications Surveys & Tutorials*, 15(3):1281–1293, Third Quarter 2013.

[14] Andy Harter, Andy Hopper, Pete Steggles, Andy Ward, and Paul Webster. The Anatomy of a Context-aware Application. In *Proceedings of the 5th Annual ACM/IEEE International Conference on Mobile Computing and Networking*, MobiCom '99, pages 59–68, Seattle, WA, USA, Aug 1999.

[15] 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 Conference*, pages 1–6, Melbourne, Australia, Nov 2005.

[16] Hiromichi Hashizume, Ayumu Kaneko, and Masanori Sugimoto. Phase Accordance Method — An Accurate Ultrasonic Positioning Method and Its Characteristics. *IEICE Transactions on Fundamentals of Electronics, Communications and Computer Sciences (Japanese Edition)*, J91-A(4):435–447, Apr 2008.

[17] Fabian Höflinger, Joan Bordoy, Nikolas Simon, Johannes Wendeberg, Leonhard Michael Reindl, and Christian Schindelhauer. Indoor-localization system for smart phones. In *2015 IEEE International Workshop on Measurements & Networking (M&N)*, pages 1–6, Coimbra, Portugal, Oct 2015.

[18] 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 *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, Sydney, Australia, Nov 2012.

[19] The International Electrotechnical Commission. IEEE Standard for a precision clock synchronization protocol for networked measurement and control systems. *IEC 61588:2009(E)*, pages C1–274, Feb 2009.

[20] The International Organization for Standardization. Acoustics — Attenuation of sound during propagation outdoors — Part 1: Calculation of the absorption of sound by the atmosphere. ISO 9613-1:1993, the International Organization for Standardization, Jun 1993.

[21] Wonho Kang and Youngnam Han. SmartPDR: Smartphone-Based Pedestrian Dead Reckoning for Indoor Localization. *IEEE Sensors Journal*, 15(5):2906–2916, May 2015.

[22] Elliott Kaplan and Christopher Hegarty. *Understanding GPS: principles and applications*. Artech house, 2005.

[23] Dilukshan Karunatilaka, Fahad Zafar, Vineetha Kalavally, and Rajendran Parthiban. LED Based Indoor Visible Light Communications: State of the Art. *IEEE Communications Surveys Tutorials*, 17(3):1649–1678, Third Quarter 2015.

[24] Kang-Wook Kim, Myung-Gon Park, Takao Hishikawa, Junghee Han, and Chang-Gun Lee. Exploiting Ultrasonic Reflections for Improving Accuracy of Indoor Location Tracking. In *2012 9th International Conference on Ubiquitous Intelligence and Computing and 9th International Conference on Autonomic and Trusted Computing*, pages 87–95, Fukuoka, Japan, Sept 2012.

[25] Yoshio Kosuge, Tadashi Koga, and Hiromi Miyazaki. Identity with TOA and TDOA Location System. *IEICE Trans. Commun. (Japanese Edition)*, J98-B(2):223–233, Feb 2015.

[26] Ye-Sheng Kuo, Pat Pannuto, Ko-Jen Hsiao, and Prabal Dutta. Luxapose: Indoor Positioning with Mobile Phones and Visible Light. In *Proceedings of the 20th Annual International Conference on Mobile Computing and Networking*, MobiCom '14, pages 447–458, Maui, HI, USA, Sept 2014.

[27] Richard B. Langley. Dilution of precision. *GPS World*, 10(5):52–59, May 1999.

[28] Gérald Lepage, Jan Bogaerts, and Guy Meynants. Time-Delay-Integration Architectures in CMOS Image Sensors. *IEEE Transactions on Electron Devices*, 56(11):2524–2533, November 2009.

[29] Binghao Li, Andrew G. Dempster, and Jian Wang. 3D DOPs for Positioning Applications Using Range Measurements. *Wireless Sensor Network*, 03(10):334–340, Oct 2011.

[30] Liqun Li, Pan Hu, Chunyi Peng, Guobin Shen, and Feng Zhao. Epsilon: A Visible Light Based Positioning System. In *Proceedings of the 11th USENIX Conference on Networked Systems Design and Implementation*, NSDI'14, pages 331–343, Seattle, WA, USA, Apr 2014.

[31] Roman Lim, Federico Ferrari, Marco Zimmerling, Christoph Walser, Philipp Sommer, and Jan Beutel. FlockLab: A Testbed for Distributed, Synchronized Tracing and Profiling of Wireless Embedded Systems. In *Proceedings of the 12th International Conference on Information Processing in Sensor Networks*, IPSN '13, pages 153–166, Philadelphia, PA, USA, Apr 2013.

[32] Roman Lim, Balz Maag, and Lothar Thiele. Time-of-Flight Aware Time Synchronization for Wireless Embedded Systems. In *Proceedings of the 2016 International Conference on Embedded Wireless Systems and Networks*, EWSN '16, pages 149–158, Graz, Austria, Feb 2016. Junction Publishing.

[33] 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, Mobicom '12, pages 305–316, Istanbul, Turkey, Aug 2012.

[34] Hui Liu and Guoqing Li. *OFDM Fundamentals*, pages 13–30. John Wiley & Sons, Inc., 2006.

[35] Kaikai Liu, Xinxin Liu, and Xiaolin Li. Guoguo: Enabling Fine-Grained Smartphone Localization via Acoustic Anchors. *IEEE Transactions on Mobile Computing*, 15(5):1144–1156, May 2016.

[36] Kaikai Liu, Xinxin Liu, Lulu Xie, and Xiaolin Li. Towards accurate acoustic localization on a smartphone. In *2013 Proceedings IEEE INFOCOM*, pages 495–499, Turin, Italy, April 2013.

[37] Laurence Mailaender. Comparing Geo-Location Bounds for TOA, TDOA, and Round-Trip Toa. In *2007 IEEE 18th International Symposium on Personal, Indoor and Mobile Radio Communications*, pages 1–5, Athens, Greece, Sept 2007.

[38] David L. Mills, Jim Martin, Jack Burbank, and William Kasch. Network Time Protocol Version 4: Protocol and Algorithms Specification. RFC 5905, June 2010.

[39] Masateru Minami, Yasuhiro Fukuju, Kazuki Hirasawa, Shigeaki Yokoyama Moriyuki Mizumachi, Hiroyuki Morikawa, and Tomonori Aoyama. DOLPHIN: A Practical Approach for Implementing a Fully Distributed Indoor Ultrasonic Positioning System. In *Proceedings of UbiComp 2004*, pages 347–365, Nottingham, UK, Sept 2004.

[40] Prasant Misra, Sanjay Jha, and Diethelm Ostry. Improving the coverage range of ultrasound-based localization systems. In *2011 IEEE Wireless Communications and Networking Conference*, pages 605–610, Cancun, Quintana Roo, Mexico, March 2011.

[41] Brian C. J. Moore. *An introduction to the psychology of hearing*. Brill, 6th ed. edition, 2013.

[42] Shotaro Murata, Chokatsu Yara, Kazumasa Kaneta, Shigenori Ioroi, and Hiroshi Tanaka. Accurate Indoor Positioning System Using Near-Ultrasonic Sound from a Smartphone. In *2014 Eighth International Conference on Next Generation Mobile Apps, Services and Technologies*, pages 13–18, Oxford, UK, Sept 2014.

[43] Shigeki Nakamura, Tomohiko Sato, Masanori Sugimoto, and Hiromichi Hashizume. An accurate technique for simultaneous measurement of 3D position and velocity of a moving object using a single ultrasonic receiver unit. In *2010 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–7, Zurich, Switzerland, Sept 2010.

[44] 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*, SenSys '07, pages 1–14, Sydney, Australia, Nov 2007.

[45] 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*, MobiCom '00, pages 32–43, Boston, MA, USA, June 2000.

[46] Nissanka B. Priyantha, Allen K. L. 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*, MobiCom '01, pages 1–14, Rome, Italy, Jul 2001.

[47] Niranjini Rajagopal, Patrick Lazik, and Anthony Rowe. Visual Light Landmarks for Mobile Devices. In *Proceedings of the 13th International Symposium on Information Processing in Sensor Networks*, IPSN '14, pages 249–260, Berlin, Germany, Apr 2014.

[48] Rafael Rodriguez, Jose A. Rabadan, Victor Guerra, Julio Rufo, and Rafael Perez-Jimenez. Distance Measurement System Based on Visible Light Communications and Ultrasound Emitters. In *ICWMC 2016, The Twelfth International Conference on Wireless and Mobile Communications*, pages 1–5, Barcelona, Spain, Nov 2016.

[49] M. M. Saad, Chris J. Bleakley, T. Ballal, and Simon Dobson. High-Accuracy Reference-Free Ultrasonic Location Estimation. *IEEE Transactions on Instrumentation and Measurement*, 61(6):1561–1570, June 2012.

[50] Shigeru Saito, Atsushi Hiyama, Tomohiro Tanikawa, and Michitaka Hirose. Indoor Marker-based Localization Using Coded Seamless Pattern for Interior Decoration. In *2007 IEEE Virtual Reality Conference*, pages 67–74, Charlotte, NC, USA, March 2007.

[51] Adam Smith, Hari Balakrishnan, Michel Goraczko, and Nissanka Priyantha. Tracking Moving Devices with the Cricket Location System. In *Proceedings of the 2nd International Conference on Mobile Systems, Applications, and Services*, MobiSys '04, pages 190–202, Boston, MA, USA, June 2004.

[52] Masanori Sugimoto, Noriyoshi Kanie, Shigeki Nakamura, and Hiromichi Hashizume. An accurate and compact 3D tracking system using a single camera and ultrasound. *Journal of Location Based Services*, 8(1):18–35, Jan 2014.

[53] Adel Thaljaoui, Thierry Val, Nejah Nasri, and Damien Brulin. BLE localization using RSSI measurements and iRingLA. In *2015 IEEE International Conference on Industrial Technology (ICIT)*, pages 2178–2183, Seville, Spain, March 2015.

[54] Andreu Urruela, Josep Sala, and Jaume Riba. Average performance analysis of circular and hyperbolic geolocation. *IEEE Transactions on Vehicular Technology*, 55(1):52–66, Jan 2006.

[55] Frank van Diggelen, Charlie Abraham, Javier de Salas, and Randy Silva. GNSS Inside Mobile Phones. *Inside GNSS*, pages 50–60, March/April 2011.

[56] Andy Ward, Alan Jones, and Andy Hopper. A new location technique for the active office. *IEEE Personal Communications*, 4(5):42–47, Oct 1997.

[57] Martin Werner, Moritz Kessel, and Chadly Marouane. Indoor positioning using smartphone camera. In *2011 International Conference on Indoor Positioning and Indoor Navigation*, pages 1–6, Guimaraes, Portugal, Sept 2011.

[58] Jiang Xiao, Zimu Zhou, Youwen Yi, and Lionel M. Ni. A Survey on Wireless Indoor Localization from the Device Perspective. *ACM Comput. Surv.*, 49(2):25:1–25:31, June 2016.

[59] Jie Yang, Simon Sidhom, Gayathri Chandrasekaran, Tam Vu, Hongbo Liu, Nicolae Cecan, Yingying Chen, Marco Gruteser, and Richard P. Martin. Detecting Driver Phone Use Leveraging Car Speakers. In *Proceedings of the 17th Annual International Conference on Mobile Computing and Networking*, MobiCom '11, pages 97–108, Las Vegas, NV, USA, Sept 2011.

Publication List

Journal papers

[1] **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Time-of-Arrival-Based Indoor Smartphone Localization Using Light-Synchronized Acoustic Waves. *IEICE TRANSACTIONS on Fundamentals of Electronics, Communications and Computer Sciences*, E100-A(9):2001–2012, 2017.

[2] Masanori Sugimoto, Hayato Kumaki, **Takayuki Akiyama**, and Hiromichi Hashizume. Optimally Modulated Illumination for Rapid and Accurate Time Synchronization. *IEEE Transactions on Signal Processing*, 65(2):505–516, 2017.

[3] Masanari Nakamura, **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Rapid and Precise Indoor 3D Localization Using Acoustic Signal for Smartphone. *IPSJ Journal*, 57(11):2489-2500, 2016.

[4] Masanari Nakamura, **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Spot Controllable Information Transfer Using Audible Frequency. *IPSJ Journal*, 57(11):2501-2514, 2016.

International conference papers

[1] **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Time-of-arrival-based Smartphone Localization Using Visible Light Communication. In *2017 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–7, Sapporo, Japan, Sept 2017.

[2] Koki Kudo, **Takayuki Akiyama**, Hiromichi Hashizume, and Masanori Sugimoto. Multicamera Synchronization for Smartphones using Optimally Modulated Illuminations. In *Proceedings of the 15th Annual International Conference on Mobile Systems, Applications, and Services* (MobiSys '17), pages 148–148, Niagara Falls, NY, USA, June 2017.

[3] **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Time Synchronization Method Using Visible Light Communication for Smartphone Localization. In *ICWMC 2016, The Twelfth International Conference on Wireless and Mobile Communications*, pages 6–9, Barcelona, Spain, Nov 2016.

[4] Masanari Nakamura, **Takayuki Akiyama**, Hiromichi Hashizume, and Masanori Sugimoto. A Spot-controllable Data Transfer Technique Using COTS Speakers. In *2016*

International Conference on Indoor Positioning and Indoor Navigation (IPIN), pages 1–8, Alcala de Henares, Spain, Oct 2016.

[5] Hayato Kumaki, **Takayuki Akiyama**, Hiromichi Hashizume, and Masanori Sugimoto. 3D Indoor Positioning and Rapid Data Transfer using Modulated Illumination. In *2016 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, Alcala de Henares, Spain, Oct 2016.

[6] **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. SyncSync: Time-of-arrival Based Localization Method Using Light-synchronized Acoustic Waves for Smartphones. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, Banff, Canada, Oct 2015.

[7] Hayato Kumaki, Masanori Sugimoto, Hiromichi Hashizume, **Takayuki Akiyama**, and Taishi Saito. A Rapid and Accurate Time-Synchronization Technique for Acoustic Localization Using Modulated Illumination. In *2015 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–10, Banff, Canada, Oct 2015.

[8] 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 of Pervasive and Ubiquitous Computing: Adjunct Publication* (UbiComp '14 Adjunct), pages 123–126, Seattle, WA, USA, Sept 2014.

[9] **Takayuki Akiyama**, Masanori Sugimoto, and Hiromichi Hashizume. Light-synchronized Acoustic ToA Measurement System for Mobile Smart Nodes. In *2014 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 749–752, Busan, Korea, Oct 2014.

[10] **Takayuki Akiyama**, Masanari Nakamura, Masanori Sugimoto, and Hiromichi Hashizume. Smart Phone Localization Method using Dual-carrier Acoustic Waves In *2013 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pages 1–9, Montbeliard-Belfort, France, Oct 2013.