Indoor Localization Using Multiple Stereo Speakers for Smartphones

Masanari Nakamura[1] and Hiroshi Kameda[1]

[1] Information Technology R&D Center, Mitsubishi Electric Corp., Kanagawa, Japan
Nakamura.Masanari@db.MitsubishiElectric.co.jp

Abstract. In this paper, we propose an acoustic indoor localization method for an embedded microphone of a smartphone using multiple stereo speakers installed indoors. In this configuration, although stereo speakers are synchronized, speakers not producing stereo sound are not synchronized. Therefore, a moving microphone could receive acoustic signals required for localization at different locations. This causes bias error in the conventional method. To address this issue, we propose a method that utilizes an asynchronous tracking filter and compensates the differences in locations of received signals using motion modelling. Through experiments, we verify that our proposed method can effectively reduce the bias error.

Keywords: Acoustic Signal, Asynchronous Tracking Filter, Smartphone, Time Difference of Arrival

1 Introduction

These days, mobile devices like smartphones are highly popular, and localization using such devices are gaining attention [1]. Although global navigation satellite system (GNSS) is a common localization method, it cannot be used indoors because receiving GNSS signals is difficult in such environments. Hence, other approaches employing the embedded sensors of smartphones are required.

Acoustic signals are suitable for accurate indoor localization using smartphones. In a related work [2-5], such signals were produced by speakers and received by a microphone; subsequently, the microphone's location was calculated. The time difference of arrival (TDoA) between the received signals provides elliptic hyperboloid, which indicates the area where the microphone exists. The estimation of location is done by calculating the intersection between multiple elliptic hyperboloids

In these methods, highly precise time synchronization of all the speakers is required. Certain audio players that can synchronize several speakers are available. However, they are more expensive than commercial off the shelf (COTS) stereo speakers, which have only two channels owing to their limited application.

Therefore, we utilize COTS stereo speakers in this study. These speakers can produce two acoustic signals simultaneously from two sides. The receiver obtains the elliptical hyperboloid from the TDoA of the speaker signals. Using the TDoAs of multiple stereo speakers, the receiver's location can be calculated. However, when the microphone is moving, TDoA can be observed at different locations, thereby rendering the intersection of elliptical hyperboloid to drift away from the true location. This is because speakers not producing stereo sound cannot produce signals simultaneously.

We herein propose an asynchronous tracking filter to compensate the difference of observation locations, and therefore reduce the bias error.

The rest of this paper is organized as follows. Chapter 2 describes the conventional method and its problem. Chapter 3 presents the proposed method for dealing with the above-mentioned problem of bias error. In Chapter 4, simulation experiments were conducted to evaluate the proposed method. Chapter 5 concludes this paper.

2 Related work

In indoor environments, some localization methods using acoustic signals utilize the embedded speakers of smartphones and microphones installed at indoor environment [6]. Whereas, other methods use smartphone's embedded microphone and speakers installed at indoor environment [2-5]. Herein, these two types of methods are referred to as active-tag system and passive-tag system, respectively. In the former, multiple signals produced by speakers of multiple smartphones collide at the microphone. Although difficult, if the transmission time of speakers is precisely synchronized, this collision is avoidable. Therefore, the latter being more preferable for localization of multiple devices, is the focus of this section.

Acoustic indoor localization methods are of two types: ranging-based and TDoA-based. The former utilizes range measurements between speakers and microphones. The latter utilizes the difference in signal received times. Ranging-based localization is typically more precise than TDoA-based localization. However, it requires high accurate time synchronization (e.g., μ order) between speakers and microphones and is therefore used for systems employing designated devices [7]. TDoA-based localization does not require such synchronizations. As accurate time synchronization is difficult in smartphones, TDoA-based localization more preferable. We herein describe the localization for a two-dimensional space for our convenience.

The relationship between the location and received time is represented as

$$\begin{cases} d_{euc}\left(\boldsymbol{r}_1 - \boldsymbol{r}(t_1)\right) - d_{euc}\left(\boldsymbol{r}_2 - \boldsymbol{r}(t_2)\right) = c \cdot (t_1 - t_2) \\ d_{euc}\left(\boldsymbol{r}_1 - \boldsymbol{r}(t_1)\right) - d_{euc}\left(\boldsymbol{r}_3 - \boldsymbol{r}(t_3)\right) = c \cdot (t_1 - t_3) \end{cases} \tag{1}$$

where the term $\boldsymbol{r}_i = (x_i, y_i)$ denotes the position vector of speaker i, and $\boldsymbol{r}(t)$ denotes the position vector of microphone at time t. t_i $(i = 1,2,3)$ denotes the received time of signal produced by speaker i, c is the sound speed, and $d_{euc}(\cdot)$ is the Euclidean distance. Equations in (1) represent hyperbolic curves of the microphone's position $\boldsymbol{r}(t)$.

When the smartphone is stationary, the positions, $\boldsymbol{r}(t_1)$, $\boldsymbol{r}(t_2)$, $\boldsymbol{r}(t_3)$ are the same and location can be calculated by solving equation (1). When the smartphone is moving, $\boldsymbol{r}(t_1)$, $\boldsymbol{r}(t_2)$, $\boldsymbol{r}(t_3)$are not the same. However, the differences are insignificant as all the speakers are synchronized. Therefore, location can be obtained without large bias error.

To simultaneously produce signals from all speakers, a special device such as a multi-channel audio player is required. As such devices are expensive, we instead use a multiple COTS stereo speaker, which hereinafter is referred to as a unit.

In this configuration, although speakers from a same unit are synchronized, those from different units are not. The relationship between the location and received time is

$$\begin{cases} d_{euc}(\boldsymbol{r}_R^1 - \boldsymbol{r}(t_R^1)) - d_{euc}(\boldsymbol{r}_L^1 - \boldsymbol{r}(t_L^1)) = c \cdot (t_R^1 - t_L^1) \\ d_{euc}(\boldsymbol{r}_R^2 - \boldsymbol{r}(t_R^2)) - d_{euc}(\boldsymbol{r}_L^2 - \boldsymbol{r}(t_L^2)) = c \cdot (t_R^2 - t_L^2) \end{cases}. \quad (2)$$

where r_R^u, r_L^u are locations of speakers R and L, belonging to unit u $(u = 1,2)$, respectively; t_R^u, t_L^u are received times of signals produced by R and L, respectively.

When the smartphone is stationary, microphone positions, $\boldsymbol{r}(t_R^1)$, $\boldsymbol{r}(t_L^1)$, $\boldsymbol{r}(t_R^2)$, and $\boldsymbol{r}(t_L^2)$ are same, and the location can be easily estimated without bias error. When the smartphone is moving, equations $\boldsymbol{r}(t_R^1) \approx \boldsymbol{r}(t_L^1)$ and $\boldsymbol{r}(t_R^2) \approx \boldsymbol{r}(t_L^2)$ are satisfied owing to speaker synchronization. However, the differences between $\boldsymbol{r}(t_R^1)$ and $\boldsymbol{r}(t_R^2)$ can increase if the smartphone starts moving. In this case, the intersection of the hyperbolic curve is drifted away from the true position (refer Figure 1-(a)), and location estimated by solving (2) incurs bias error.

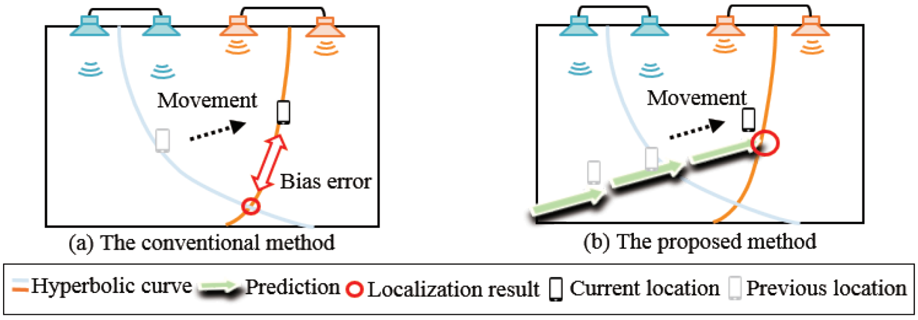


Fig. 1. Localization of moving smartphone using multiple stereo speakers

Takabayashi [8] proposed a bias error reduction method for single active-tag systems. We herein apply this method to a passive-tag system with multiple COTS stereo speakers.

3 Proposed method

3.1 System architecture

In the proposed method, two speakers of a unit simultaneously produce acoustic signals. If the transmission time of each unit is not controlled, numerous signals could collide with each other at the microphone, thereby resulting in large errors. Therefore, the transmission time of each unit should be controlled using general communication systems such as Wi-Fi and Bluetooth.

Speaker time lags in conventional methods produce bias errors, therefore necessitating a multiple-channel audio player to synchronize all speakers [2].

The smartphone's current position is estimated using an asynchronous tracking filter when a TDoA is measured. The estimation is done by calculating the intersection between the predicted position and hyperbolic curve of TDoA, as shown in Figure 1-(b), and it reduces the bias error as shown in Figure 1-(a), the details of which are described in the next section.

3.2 Asynchronous tracking filter.

TDoA observation model and motion model.

The state vector, $\boldsymbol{x}$ comprises position and velocity $[x \quad y \quad \dot{x} \quad \dot{y}]^{tr}$ where x, y are positions, $\dot{x}, \dot{y}$ are velocities, and tr is the transpose. The relationship between the state and position is represented as

$$\boldsymbol{r} = \begin{bmatrix} 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 \end{bmatrix} \boldsymbol{x} = \boldsymbol{H}\boldsymbol{x} \tag{3}$$

In the asynchronous tracking filter, the estimation is conducted sequentially at every TDoA measurement, comprising t_R^u and t_L^u. We herein define the lesser of t_R^u or t_L^u as the TDoA observation time. t_k denotes as the observation time of k-th TDoA.

The TDoA measurement, z_k^u received at t_k, corresponding to unit u is represented as

$$z_k^u = \left(t_R^u + n_{t_R^u}\right) - \left(t_L^u + n_{t_L^u}\right) = h^u(\boldsymbol{x_k}) + \left(n_{t_R^u} - n_{t_L^u}\right) \tag{4}$$

where $n_{t_R^u}, n_{t_L^u}$ are the measured noises of received times produced by R and L of the unit u, respectively. $\boldsymbol{x_k}$ is the state vector $[x_k \quad y_k \quad \dot{x}_k \quad \dot{y}_k]^{tr}$ at time t_k. From equation (2), $h^u(\boldsymbol{x_k})$ represents

$$h^u(\boldsymbol{x}_k) = \frac{d_{euc}(\boldsymbol{r}_R^u - \boldsymbol{H}\boldsymbol{x}_k) - d_{euc}(\boldsymbol{r}_L^u - \boldsymbol{H}\boldsymbol{x}_k)}{c} \tag{5}$$

The observation noises $n_{t_R^u}, n_{t_L^u}$ are white Gaussian noise with zero-mean and known variance $\sigma_{t_R^u}^2$, $\sigma_{t_L^u}^2$, respectively. The measurement error of TDoA n_k^u is $n_{t_R^u} - n_{t_L^u}$. Notably, we assume that $n_{t_R^u}$ *and* $n_{t_L^u}$ are independent. In this case, n_k^u is white Gaussian noise with zero-mean and variance $\sigma_{t_R^u}^2 + \sigma_{t_L^u}^2$ because Gaussian distribution has the reproductive property.

As the motion model of the microphone, we utilize the constant velocity model.

$$\boldsymbol{x}_k = \begin{bmatrix} 1 & 0 & \Delta t_k & 0 \\ 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & \Delta t_k \\ 0 & 0 & 0 & 1 \end{bmatrix} \boldsymbol{x}_{k-1} + \begin{bmatrix} \Delta t_k & 0 \\ 0 & \Delta t_k \\ 1 & 0 \\ 0 & 1 \end{bmatrix} \boldsymbol{w}_k = \boldsymbol{F}(\Delta t_k)\boldsymbol{x}_{k-1} + \boldsymbol{\Gamma}\boldsymbol{w}_k \tag{6}$$

where $\boldsymbol{w}_{\mathbf{k}}$ is the process noise vector $\left[w_k^x \quad w_k^y\right]^{tr}$ which represents ambiguity of the motion. w_k^x and w_k^y are white Gaussian noise with zero-mean and variance σ_w^2. Δt_k is the difference of the observation times $t_k - t_{k-1}$.

Algorithm

We herein describe the algorithm to estimate the microphone state. When the first measurement is inputted to the asynchronous tracking filter ($k = 1$), particles $\boldsymbol{x}_k^j$ representing the microphone state are generated ($j = 1, \ldots, J$). Each element of the state $\boldsymbol{x}_k^j = \left[x_k^j \quad y_k^j \quad \dot{x}^j{}_k \quad \dot{y}^j{}_k\right]^{tr}$ is generated by uniform random numbers. The weight α_k^j of the state $\boldsymbol{x}_k^j$ is given as $\alpha_k^j = 1/J$. When $\boldsymbol{x}_k^j$ and α_k^j were already generated ($k \geq 2$), the prediction based on the motion model (6) is conducted for each particle.

In the updating step, the likelihood of each particle is calculated using z_k^u and the weight α_k^j is updated.

$$\alpha_k^j = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{\left(z_k^u - h^u(\boldsymbol{x}_k^j)\right)^2}{2\sigma^2}\right) \tag{7}$$

The estimated state $\hat{\boldsymbol{x}}_k$ is obtained by the weighted sum $\sum_{j=1}^{J} \alpha_k^j \cdot \boldsymbol{x}_k^j$.

When the next measurement is inputted, the above-mentioned process is repeated with the resampled particles [9].

4 Numerical result

4.1 Scenarios

We conducted following simulation experiments for the evaluation of our proposed method. Figure 2 shows the microphone route and the four speaker locations. The microphone moves at a speed of 1 m/sec and receives its first TDoA measurement at the route's starting position (-3.0, 1.5).

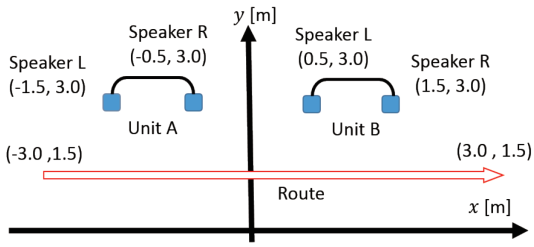


Fig. 2. Simulation scenario

At indoor environments, the transmission interval is designed according to the signal length and the reverberation time of multipath. Regarding the signal length, a short one with large amplitude is preferable for the precision and the updating rate of localization. However, the amplitude is generally restricted due to speaker's inaudibility. To obtain the desired signal-to-noise ratio (SNR) with the restricted amplitude, a longer signal can be utilized. A reverberation time of 100 ms is sufficient to attenuate the multipath [4]. Considering the above-mentioned reasons, transmission intervals of these simulations are set to 100 ms and 300 ms for the short and the long signal cases, respectively.

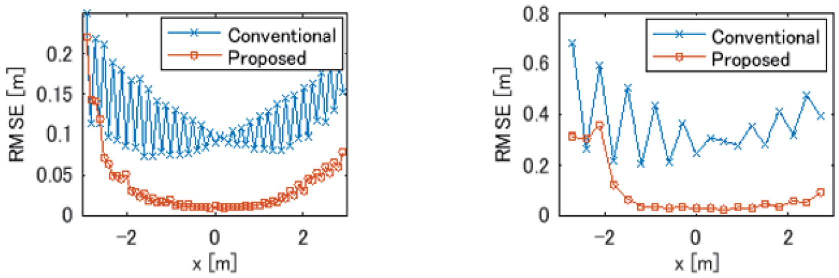
The observation noise of the received time σ depends on the bandwidth and SNR. The bandwidth and SNR herein are set to 1 kHz and 30 dB, respectively. These parameters were determined in consideration of the characteristic of microphones embedded in smartphones [4-5]. In this case, the observation noise σ is 1.58×10^{-5} sec. Hence, the observation noise of TDoA is $2 \times 1.58 \times 10^{-5} = 3.16 \times 10^{-5}$ sec.

The conventional method numerically solves equation (2) using the current and previous measurements. The observation noise, process noise, and number of particles of the proposed method were set to 3.16×10^{-5} sec, 0.5 m/sec, and 5000, respectively. The range of uniform random numbers to generate these particles $\left[x_k^j \quad y_k^j \quad \dot{x}^j{}_k \quad \dot{y}^j{}_k\right]'$ at $k = 1$ were [-3, 3], [-1, 3], [-1.5, 1.5], [-1.5, 1.5] m, respectively.

Root mean square error (RMSE) was used for evaluating each trial with the number of trials set to 100. The starting position $k = 1$ was excluded from this evaluation as the conventional method cannot estimate it.

4.2 Results and Discussion

Figures 3-(a) and 3-(b) show the results of conventional and proposed methods with transmission intervals set to 100 ms and 300 ms, respectively. The horizontal axis shows the signal receiving position, clearly indicating the relationship between the RMSE and the receiving position as well as effectiveness of the proposed method in RMSE reduction.



(a) Transmission Interval 100ms (b) Transmission Interval 300ms

Fig. 3. RMSE of evaluation result

Figures 4-(a) and 4-(b) show examples of results obtained by the conventional and proposed methods, respectively. In Figure 4-(a), estimated locations were around the intersections of hyperbolic curves, far from true positions, which represent the bias errors. Figure 4-(b) indicates that the proposed method can reduce these errors using the asynchronous tracking filter.

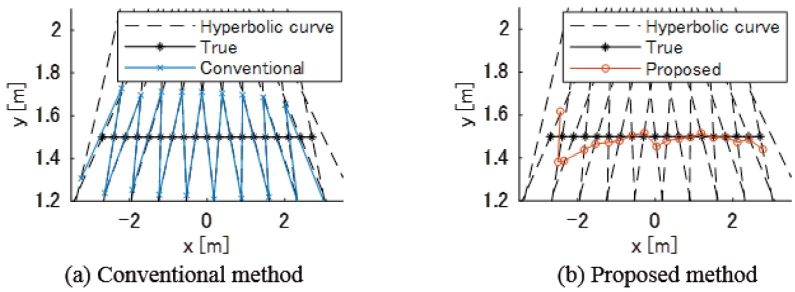


Fig. 4. Example of estimated locations (transmission interval: 300 ms)

In Figures 3-(a) and 3-(b), RMSE of the proposed method was minimum around $x = 0$ m, turning slightly worse in $x > 0$ m. For clarification, predicted particles at $x = 0$ m and 2.5 m were plotted in Figures 5-(a) and 5-(b). The transmission interval was set to 300 ms. In these figures, the particles are scattered in an ellipse, and the major axis was roughly parallel to the previously plotted hyperbolic curve. This is because these particles were generated by extrapolating the resampled particles based on the previously obtained hyperbolic curve.

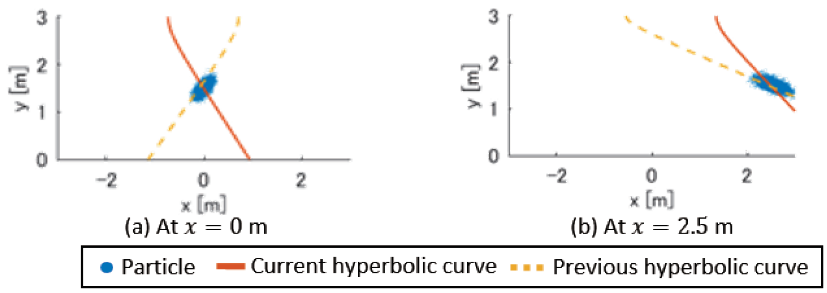


Fig. 5. Predicted particles distribution

When the microphone's position was $x = 0$ m (Figure 5-(a)), current hyperbolic curve crossed the major axis of the ellipse nearly at right angles. In this case, the area of resampled particles was narrow, leading to high precision. When the microphone's position was $x = 2.5$ m (Figure 5-(b)), the minor axis of the ellipse crossed the current hyperbolic curve at roughly right angles, and the area of resampled particles was relatively broad, leading to reduction in precision.

5 Conclusion

In this paper, we described an acoustic indoor localization system with multiple COTS stereo speakers. We proposed the asynchronous tracking filter to reduce the bias error caused by the asynchrony of speakers that do not produce stereo sound when the smartphone is moving. The simulation experiments showed that the proposed method can effectively reduce the bias error.

References

1. J. Xiao, Z. Zhou, Y. Yi and L. M. Ni, "A survey on wireless indoor localization from the device perspective," *ACM Computing Surveys,* vol. 49, no. 2, pp. 25:1-25:31, (2016).
2. K. Liu, X. Liu and X. Li, "Guoguo: Enabling Fine-Grained Smartphone Localization via Acoustic Anchors," *IEEE Trans. on Mobile Computing,* vol. 15, no. 5, pp. 1144-1156, (2016).
3. F. J. Álvarez, T. Aguilera and R. L. Valcarce, "CDMA-based acoustic local positioning system for portable devices with multipath cancellation," *Digital Signal Processing,* vol. 62, pp. 38-51, (2017).
4. T. Akiyama, M. Nakamura, M. Sugimoto and H. Hashizume, "Smart phone localization method using dual-carrier acoustic waves," *Proc. of IPIN 2013,* pp. 1-9, 2013.
5. M. Nakamura, T. Akiyama, M. Sugimoto and H. Hashizume, "3D FDM-PAM: rapid and precise indoor 3D localization using acoustic signal for smartphone," *Proc. of ACM Ubicomp 2014,* pp. 123-126, (2014).
6. F. Höflinger, R. Zhang, J. Hoppe, A. Bannoura, A. Reindl, J. Wendeberg, M. Buhrer and C. Schindelhauer, "Acoustic Self-calibrating System for Indoor Smartphone Tracking (ASSIST)," *Proc. of IEEE IPIN 2012,* pp. 1-9, (2012).
7. N. Priyantha, A. Chakraborty and H. Balakrishnan, 'The Cricket Location Support System," *Proc. of ACM MobiCom 2000,* pp. 32-43, (2000).
8. Y. Takabayashi, T. Matsuzaki, H. Kameda and M. Ito, "Target tracking using TDOA/FDOA measurements in the distributed sensor network," *Proc. of SICE 2008,* pp. 3441-3446, (2008).
9. S. M. Arulampalam, S. Maskell, N. Gordon and T. Clapp, "A Tutorial on Particle Filters for Online Nonlinear/Non-Gaussian Bayesian Tracking," *IEEE Trans. on Signal Processing,* vol. 50, no. 2, pp. 174-188, (2002).
10. R. A. Mark, J. A. Scheer and W. A. Holm, "Principles of Modern Radar: Basic Principles," SciTech Publishing, (2010).