## Article

# Indoor Pedestrian Self-Positioning Based on Image Acoustic Source Impulse Using a Sensor-Rich Smartphone 

Xiyu Song ${ }^{1(1)}$, Mei Wang ${ }^{2}$, Hongbing Qiu ${ }^{1, *}$ and Liyan Luo ${ }^{3}$<br>1 Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, GuiLin University of Electronic Technology, Guilin 541004, China; songxiyu@guet.edu.cn<br>2 College of Information Science and Engineering, Guilin University of Technology, Guilin 541004, China; mwang@guet.edu.cn<br>3 Wireless Broadband and Signal Processing Guangxi Key Laboratory, Guilin University of Electronic Technology, Guilin 541004, China; xiaoyan12027@gmail.com<br>* Correspondence: qiuhb@guet.edu.cn; Tel.: +86-159-7734-2353

Received: 24 September 2018; Accepted: 22 November 2018; Published: 26 November 2018


#### Abstract

The ubiquity of sensor-rich smartphones provides opportunities for a low-cost method to track indoor pedestrians. In this situation, pedestrian dead reckoning (PDR) is a widely used technology; however, its cumulative error seriously affects its accuracy. This paper presents a method of combining infrastructure-free indoor acoustic self-positioning with PDR self-positioning, which verifies the rationality of PDR results through the acoustic constraint between a sound source and its image sources. We further determine the first-order echo delay measurements, thus obtaining the mobile user position. We verify that the proposed method can achieve a continuous self-positioning median error of 0.19 m , and the error probability below 0.12 m is $54.46 \%$, which indicates its ability to eliminate PDR error, as well as its adaptability to environmental disturbances.


Keywords: indoor acoustic localization; acoustic image model; motion dynamics information; pedestrian dead reckoning; smartphone-based self-positioning

## 1. Introduction

The increasing number of sensor-rich smartphones has raised interest in using their sensors for indoor localization applications, such as indoor navigation [1], location-based services [2], providing aid for hearing-impaired persons [3], and environmental perception [4,5]. Global positioning system (GPS) can provide effective localization results for pedestrians in outdoor environments, but may not be useful for indoor environments due to weak signal reception and the indoor shadowing effect [6]. Therefore, indoor pedestrian self-positioning technology has attracted considerable attention.

Based on specific technology, it is possible to categorize methodologies for smartphone-based indoor pedestrian self-positioning systems into two distinct groups: (1) infrastructure-based systems that use auxiliary equipment or a cooperation between nodes to realize target tracking [1-9], and (2) the infrastructure-free systems that realize pedestrian self-positioning using only the information provided by the smartphone carried on one's person [9-20]. However, when using the former, the pedestrian is likely to experience difficulties in position acquisition when the cooperative information is unavailable. The widely used pedestrian dead reckoning (PDR) method only provides a relative position estimate, with its accuracy degrading over time. The fusion of other positioning methods has been proposed to solve this problem [21-26]. Yang et al. [23] proposed a novel smartphone-based indoor localization system that improved the PDR results by integrating an infrastructure-based acoustic localization
system, reaching sub-meter localization accuracy at the expense of a complicated data availability analysis and computational complexity.

To reduce the influences of noise on the source tracking, the motion and observation models of the moving source as well as the probability distribution model of the errors [27] must be established for filtering methods, i.e., Kalman filter, particle filter, and their variants [26]. Such complications and inconveniences limit the applications of the filtering methods.

To alleviate the aforementioned problems, this study provides an acoustic constraint algorithm to verify the rationality of the PDR results, which reduces the cumulative errors by using the geometric relationship between the sound source and its image sources.

The rest of this paper is organized as follows. Section 2 provides an overview of the proposed indoor pedestrian self-positioning system. Section 3 details the first-order echo estimates based on an acoustic image model (AIM). In Section 4, we describe the solution for the first-order echo measurements in three steps: the calculation of the cross-correlation, the calculation of first-order echo measurements; and the acoustic principle-based constraints. Section 5 summarizes the applied Levenberg-Marquardt algorithm-based weighted nonlinear least squares (LMA-WNLS) model for pedestrian position values. Section 6 highlights the performance of the proposed method and the results analysis, which proves the effectiveness of the proposed system for indoor pedestrian continuous position acquisition. The conclusions are drawn in Section 7.

## 2. System Overview

We assume that a sounding smartphone is always carried by the indoor pedestrian. The pedestrian moves autonomously inside a room. At every step, the loudspeaker of the smartphone produces a chirp pulse, the microphone of this smartphone registers the echoes, and the inertial sensors record the accelerometer and gyroscope readings. We define the room to be a K-faced rectangular room, which is widely used in teaching buildings. The pedestrian is modeled in a room as a point source in a rectangular cavity, and thus, for ease of explanation, the pedestrian and the sound source (the loudspeaker of the user smartphone) are hereafter equivalently used in this paper. We worked in two-dimensional (2D) space, ignoring the floor and the ceiling, given $K=4$, but the results could be extended to three-dimensional space. The proposed system was implemented to achieve submeter-level positioning accuracy and reliability. To this end, five steps were followed to obtain the position of the indoor pedestrian, as presented in Figure 1.


Figure 1. Overview of the proposed system architecture. AIM: Acoustic Image Model. ITM: Isosceles Trapezoid Model; LMA-WNLS: Levenberg-Marquardt algorithm-based weighted nonlinear least squares.

The first step is to compute the image sound sources, denoted as $S_{t, k}, k=1, \ldots, 4$ as shown in Figure 2, without loss of generality. One corner marked with $O$ in the room is placed as the origin based on the AIM $[28,29]$. Then, Euclidean distance analysis is applied for the first-order echo estimates, which are detailed in Section 3. An isosceles trapezoid geometry [20] was adopted to calculate the first-order echo measurements based on the PDR information (i.e., the step length $L_{\text {tra }}$ [30] and the heading angular $\theta$ [31]) and the locations of all $S_{t, k}$. The fourth step is to exploit the acoustic constraits to update the measurementsvalues. Lastly, the LMA-WNLS is performed, which is used to quickly iterate the current pedestrian position coordinates and achieve the tracking effect. The LMA-WNLS is detailed by Mensing [32], and a brief summary is provided in Section 5.


Figure 2. Illustration of the spatial geometrical models presented in this paper. We suppose the room size is $\left[L_{x}, L_{y}\right]$, and the origin point is located at $O(0,0)$. (a) The acoustic image model (AIM) for the first-order images. Point $p$ is an arbitrary point of the $k$ th wall. Vector $n$ is the outward-pointing unit normal associated with the $k$ th wall, $S_{t, k}, k=1, \ldots, K$ are the first-order image sources of $S_{t}$ corresponding to the $k$ th wall. (b) The isosceles trapezoid model (ITM) for a moving sound source. When the sound source moves from $S_{t}$ to $S_{t+1}, S_{t, k}$ moves to the $S_{t+1, k}, k=1, \ldots, 4$, respectively; then, these points ( $S_{t}, S_{t, k}, S_{t+1}, S_{t+1, k}$ ) can form a set of isosceles trapezoids with the waist length represented as the step length $L_{\text {tra }}$ and the inner angle as the heading angular $\theta$. The step forward from $S_{t}$ to $S_{t+1}$ is shown as a green full line, the sound rays at time $t$ are the blue dashed line, and sound rays at time $t+1$ are the red dotted line.

## 3. First-Order Echo Estimates

In the AIM, the reflections from the walls are replaced with signals produced by image sound sources across the corresponding walls. For a first-order echo and the $k$ th wall described by the outward-pointing unit normal $n_{k}$ and an arbitrary wall point $p_{k}$, the image sources $S_{t, k}$ of the real source $S_{t}$ are computed as:

$$
\begin{equation*}
S_{t, k}=S_{t}+2\left\langle p_{k}-S_{t}, n_{k}\right\rangle n_{k} \tag{1}
\end{equation*}
$$

where $\langle\cdot\rangle$ is the inner product operator. According to Equation (1), given $n_{k}$ and $p_{k}, S_{t, k}$ can be determined by $S_{t}$ using the dimension analysis introduced by Figure 1 in Fu et al. [29]. For example, when $k=3$ (the east wall), the unit normal $n_{3}=(1,0)$ and the wall point $p_{3}=\left(L_{x}, 0\right)$. Supposing a real sound source is located at $S_{t}(x, y)$, its first-order image sound sources are located at $S_{t, k}\left(x_{t, k}, y_{t, k}\right)$ for the $k$ th wall at time $t$, and then $S_{t, 3}\left(x_{t, 3}, y_{t, 3}\right)$ could be $S_{t, 3}\left(2 L_{x}-x, y\right)$ using Equation (1). Similarly, the other images' positions could be computed as shown in Table 1.

Table 1. Suppose a real sound source is located at $S_{t}(x, y)$ : its first-order image sound sources are located at $S_{t, k}\left(x_{t, k}, y_{t, k}\right)$ for different $k$ at any time $t$. The corresponding coordinates and reflection orders are shown below.

| Coordinate | $\boldsymbol{- 1 s t}$ Order | Real Source | 1st Order |
| :---: | :---: | :---: | :---: |
| $x$ | $S_{t, 1}(-x, y)$ | $S_{t}(x, y)$ | $S_{t, 3}\left(2 L_{x}-x, y\right)$ |
| $y$ | $S_{t, 2}(x,-y)$ | $S_{t}(x, y)$ | $S_{t, 4}\left(x, 2 L_{y}-y\right)$ |

Denote $r_{e t, k}$ as the Euclidean distance between the $S_{t}$ and its $S_{t, k}$ at time $t$, then:

$$
\begin{equation*}
r_{e t, k}=\left\|S_{t}-S_{t, k}\right\|=\sqrt{\left(x-x_{t, k}\right)^{2}+\left(y-y_{t, k}\right)^{2}}, k=1,2, \ldots, K \tag{2}
\end{equation*}
$$

As the sound propagation speed $c$ is used as a constant here, in the following, we treat distances and propagation times as equivalent. Thus, the first-order echo estimates $d\left(r_{e t, k}\right)$ as a delay set for real sound source at time $t$ could be expressed by the difference of $r_{e t, k}$ as:

$$
\begin{equation*}
d\left(r_{e t, k}\right)=\left[r_{e t, 2}-r_{e t, 1}, r_{e t, 3}-r_{e t, 1}, \ldots, r_{e_{t, \hat{k}}}-r_{e t, 1}\right], \hat{k}=2,3, \ldots, K \tag{3}
\end{equation*}
$$

## 4. First-Order Echo Measurements

When the loudspeaker $S$ of the smartphone chirps in an indoor environment, the smartphone microphone $M$ records both the direct path of the sound and its reflections from the walls. Motivated by the robustness of the transfer-function measurement approach based on sequences with better cross-correlation and autocorrelation properties [33], a chirp impulse [16] with similar properties and those more compatible with smartphones [7] was chosen as the emitting signal to simplify the processing of the first-order echo measurements using the generalized cross correlation (GCC) introduced by Knapp et al. [34], which performed well in separating arrivals that were close in time.

### 4.1. Calculation of the Cross-Correlation

The chirp impulse, emitted from $S$, works between $0 \leq t \leq T$ with a start frequency $f_{0}$ and an end frequency $f_{1}$, which can be described as:

$$
\begin{equation*}
s(t)=\sin \left(2 \pi\left(f_{0}+\frac{f_{1}-f_{0}}{2 T} t\right) t\right) \tag{4}
\end{equation*}
$$

Let the time-domain received signals be $r(t)$, the GCC between the received signals $r(t)$ and the reference signal $s(t)$ is given by the phase transform (PHAT) in time domain:

$$
\begin{equation*}
R(\tau)_{s, r}=\frac{1}{2 \pi} \int_{-\infty}^{+\infty} \frac{s(\omega) r^{*}(\omega)}{|s(\omega) r(\omega)|} e^{j \omega \tau} d \omega \tag{5}
\end{equation*}
$$

where $*$ is the conjugate operator, $s(\omega)$ and $r(\omega)$ represent the Fourier transforms of the reference signal and the signals received by the microphone of the smartphone, respectively. The GCC-PHAT method has several advantages: first, the correlation between the received signals with a known signal removes uncorrelated noise; second, the implementation of the cross correlation in the frequency domain is more computationally efficient than its implementation in the time domain; third, the PHAT has the ability to decrease the effects of reverberation [35]. In our experiments, since the pedestrian walks along the room's walls and the dominant directions (east, west, south, north), shown by the reference walking lines in Figure 3, the distance to the four walls are not always equal and the range-resolution is sufficient for path separation. Thus, given the advantage of the chirp's good correlation characteristics, the GCC-PHAT $R(\tau)_{s, r}$ has the ability to detect the time-of-flights (TOFs), both of the direct path and the reflected path.


Figure 3. Illustration of the fifth corridor of the Jinji Campus Library in GUET. GUET: Guilin University of Electronic Technology. The dashed lines are the reference walking lines. A small green triangle dot denotes the beginning point and a red one denotes the ending point. The dominant directions are denoted as E: East, S: South, W: West, N: North.

### 4.2. Calculation of First-Order Echo Measurements

Given a fixed reflecting surface with a fixed orientation and a sound source point, the expression for the position of the image point can obtained with Equation (1). If also given the boundary values of the room size, this position can be explicitly expressed by Table 1. Thus, according to Equation (2), the distance relationship between the real sound source and its first-order image sound sources can be expressed by taking advantage of the isosceles trapezoid model (ITM), shown in Figure 2b, as:

$$
\left\{\begin{array}{l}
r_{m t+1, k}=r_{m t, 1} \pm 2 L_{t r a} \cos \theta, k=1,3, S_{t+1, x}>S_{t, x}  \tag{6}\\
r_{m t+1, k}=r_{m t, 1} \pm 2 L_{t r a} \sin \theta, k=2,4, S_{t+1, y}>S_{t, y} \\
r_{m t+1, k}=r_{m t, 1} \mp 2 L_{t r a} \cos \theta, k=1,3, S_{t+1, x}<S_{t, x} \\
r_{m t+1, k}=r_{m t, 1} \mp 2 L_{t r a} \sin \theta, k=2,4, S_{t+1, y}<S_{t, y}
\end{array}\right.
$$

where $r_{m t, k}$ is the distance between $S_{t, k}$ and $S_{t+1, k}$ for the $k t h$ wall at time $t$. In the aforementioned expression, the dependence on the wall index $k$ is omitted for the sake of brevity; here, $k=1$ is specifically the west wall. $r_{m t+1, k}$ is the distance for the $k t h$ wall at time $t+1 . S_{t}$ and $S_{t+1}$ with the subscript $x$ or $y$ are the corresponding coordinate values of $S$ at time $t$ and time $t+1$, respectively.

Since the smartphone is carried by the moving pedestrian, the PDR information-which is regarded as the distance moved (the step length $L_{\text {tra }}$ ), and the movement heading attitude changes (the heading angle $\theta$ )—could be solved by the adaptive step length algorithm presented by Shin et al. [30] and a heading correction method similar to the one presented by Deng et al [31]. Denoting WF as the walking frequency when the steps are detected and $A V$ as the acceleration variance, the step length is a linear function of the following measurements:

$$
\begin{equation*}
L_{t r a}=\left(m_{w f} \alpha_{w f}\right) W F+\left(m_{a v} \alpha_{a v}\right) A V+\left(v_{w f}+v_{a v}\right)=\alpha_{O p t} W F+\beta_{O p t} A V+\gamma_{O p t} \tag{7}
\end{equation*}
$$

where $m_{w f}$ and $m_{a v}$ are the measurement errors of $W F$ and $A V$, respectively, and they are both equal to 0.5 , because the measurement errors are minimized. $\alpha_{w f}$ and $v_{w f}$, as well as $\alpha_{a v}$ and $v_{a v}$, are the linear fit parameters for $W F$ and $A V$, respectively. In our experiment, the parameters of $\alpha_{w f}$ and $v_{w f}$, as well as $\alpha_{a v}$ and $v_{a v}$, were obtained by averaging the results by recording multiple measurements on the same experimental route. Thus, $\alpha_{O p t}, \beta_{O p t}$, and $\gamma_{O p t}$ are the optimal step length estimation parameters.

As the real paths of the experimenter in this study were along the dominant directions, and during the experiment, the smartphone was always horizontally and statically held in the hand, we simplified the processing of $\theta$ by superimposing the $z$-axis angular rate reading $\omega_{z, \kappa}$ from the gyroscope at every step $\kappa .0 \leq \kappa \leq \mathcal{K}$, where $\mathcal{K}$ is the total step number:

$$
\begin{equation*}
\theta=\sum_{\kappa=1}^{\mathcal{K}} \omega_{z, \kappa} \tag{8}
\end{equation*}
$$

Similarly, the first-order echo measurements $d\left(r_{m t, k}\right)$ as a delay set for $S$ at time $t$ could be

$$
\begin{equation*}
d\left(r_{m t, k}\right)=\left[r_{m t, 2}-r_{m t, 1}, r_{m t, 3}-r_{m t, 1}, \ldots, r_{m_{t}, \hat{k}}-r_{m t, 1}\right], \hat{k}=2,3, \ldots, K \tag{9}
\end{equation*}
$$

### 4.3. Acoustic Principle-Based Constraints

Since the distance between $S$ and $M$ is very small, i.e., $\|S-M\| \rightarrow 0$, the direct sound path from $S$ to $M$ can be described as:

$$
\begin{equation*}
\tau_{\text {direct }}=\underset{\tau}{\operatorname{argmax}}\left|R(\tau)_{s, r}\right| \tag{10}
\end{equation*}
$$

where $|\cdot|$ is the modulo operation. Similarly, the direct sound paths from $S_{t, k}$ to $M$ can be regarded as the path from $S_{t, k}$ to $S$, and we denote $\tau_{\text {reflect }}$ as a TOF set of these paths as:

$$
\begin{equation*}
\tau_{\text {reflect }}=\underset{\tau}{\arg }\left(\left|R(\tau)_{s, r}\right|-\max \left|R(\tau)_{s, r}\right|\right) \tag{11}
\end{equation*}
$$

The first-order echo measurements are provided by solving the unknown top or base values of the isosceles trapezoids that should be the impulse delays in the $R(\tau)_{s, r}$. To reduce the errors of the first-order echo measurements, the acoustic principle-based constraint algorithm is proposed to update the measurements.

### 4.3.1. Sound Pressure Level Constraint

The Haas effect, also known as the priority effect, reflects the perception of the sound source's orientation based on the first sound that arrives at the human ear. According to the conclusion of the classic Hass experiment, sounds reflected within 5 to 35 ms after the direct sound can be distinguished when the sound pressure level (SPL) of the reflected sound is greater than 10 dB of the SPL of the direct sound. Thus:

$$
\begin{equation*}
S P L_{\text {reflect }}-S P L_{\text {direct }} \geq 10 \mathrm{~dB} \tag{12}
\end{equation*}
$$

where $S P L_{\text {reflect }}$ and $S P L_{\text {direct }}$ are the SPLs of the first-order reflections and the direct arrived sound, respectively. Since the sound source is a point source, assume that the image sources are also point sources, so the wavefront is a spherical wave. The expression of spherical acoustic wave attenuation with distance at normal temperature is:

$$
\begin{equation*}
S P L=L W-10 \lg r-k \tag{13}
\end{equation*}
$$

where the $L W$ is the sound power level, $r$ is the distance between the sound source (the real source or the image source) and the receiver, and $k$ is the spacial modifying coefficient. Let $L W_{\text {reflect }}$ and $L W_{\text {direct }}$ be the sound power level at the real sound source and its first-order image sound source, respectively. Based on the image concepts in AIM, $L W_{\text {reflect }}=L W_{\text {direct. }}$. Then:

$$
\begin{align*}
S P L_{r e f l e c t}- & S P L_{\text {direct }} \\
& =L W_{\text {reflect }}-20 \lg r_{\text {reflect }}-k-\left(L W_{\text {direct }}-20 \lg r_{\text {direct }}-k\right)  \tag{14}\\
& =20 \lg r_{\text {direct }}-20 \lg r_{\text {reflect }}=20 \lg \left(\frac{c \tau_{\text {direct }}}{c \tau_{\text {reflect }}}\right)
\end{align*}
$$

Thus, $\tau_{\text {reflect }} \leq \frac{\tau_{\text {direct }}}{\sqrt{10}}$, which means if any first-order reflected sounds within 5 to 35 ms after the direct sound, the $r_{m t, k}, t \geq 0$ must follow:

$$
\begin{equation*}
r_{m t, k} \leq \frac{c \cdot \tau_{\text {direct }}}{\sqrt{10}} \tag{15}
\end{equation*}
$$

If some of them (the $r_{m t, k}$ ) are outside this range, the known room size $\left[L_{x}, L_{y}\right]$ should be used to restrict their values. For example, when the pedestrian walks along the west wall $(k=1), r_{m t, 1}$ should be the smallest one among all the $r_{m t, k}$ values along the west wall phase, $r_{m t, 1}$ must follow Equation (15); however, the first-order echo delay according to the opposite side (the east wall, $k=3$ ) in this phase may be outside the 5 to 35 ms range, then the $r_{m t, 3}$ value should be restricted by $L_{x}$, i.e., $r_{m t, 3}=L_{x}-r_{m t, 1}$. A similar analysis also applies to the $r_{m t, 2}$ and $r_{m t, 4}$ with $L_{y}$.

### 4.3.2. Sound Energy Constraint

Based on the distance relationship between the real sound source and its image sound sources, the propagation delay $\xi_{t, k}$ for any $r_{m t, k}, t \geq 0$ is:

$$
\begin{equation*}
\xi_{t, k}=\operatorname{round}\left(\frac{r_{m t, k}}{c}\right) \tag{16}
\end{equation*}
$$

where $\operatorname{round}(\cdot)$ is the rounding operation. As $\xi_{t, k}$ should be a TOF value in $R(\tau)_{s, r}$, and the computed $\frac{r_{m t, k}}{c}$ is not always an integer, a rounding operation is needed. Based on the fact that the energy of the wave is proportional to the square of its amplitude, the pulse amplitude of the cross-correlation function could be used to represent the energy constraint. The sound energy (SE) constraint of the first-order echo impulses according to $\xi_{t, k}$ should be:

$$
\begin{equation*}
\left|R\left(\xi_{t, k}\right)_{s, r}\right|^{2} \geq \Delta \tag{17}
\end{equation*}
$$

where $\Delta$ is an empirical energy threshold that depends on the room average absorption coefficient. Because the four sides of our experimental environment are glass windows, doors and walls, and the ceiling is mainly glass with steel stent supports (as shown in Figure 3), according to the sound absorption coefficient analysis [36], the sound field is not uniform. Under these conditions, the calculated coefficient will always be smaller than when the sound field is uniform. We calculate the indoor reverberation time according to the Sabine formula, confirming that the room is a high reverberation environment. This may result in the superposition of multiple reflected sounds at the position where the first-order reflected wave occurs. In addition, in large rooms, the sound propagation will experience a long path, when the frequency is above 2 kHz , the air absorption can account for $20-25 \%$ of the total sound absorption of the whole space. Therefore, through experimental observation, our empirical energy threshold is set as the following

$$
\begin{equation*}
\Delta=\left(\frac{1}{2 \mathcal{K}} \sum_{i=1}^{\mathcal{K}} \max \left|R_{i}\left(\tau_{\text {reflect }}\right)_{s, r}\right|\right)^{2} \tag{18}
\end{equation*}
$$

### 4.3.3. Update Algorithm

If the first-order echo measurements $d\left(r_{m t, k}\right)$ satisfy the SPL and SE constraints, meaning the PDR is authentic, $d\left(r_{m t, k}\right)$ is correct. If not, the PDR is not completely authentic, and $d\left(r_{m t, k}\right)$ should be updated by the new values extracted from the constraint range. The above constraint steps are summarized in Algorithm 1.

```
Algorithm 1. Algorithm for Updated First-Order Echo Measurements.
    Input: \(r_{m t, k}, \xi_{t, k}, \tau_{\text {direct }}, c, R(\tau)_{s, r}, \Delta, \tau_{\text {reflect }}\)
    Output: the updated \(d\left(r_{m t, k}\right)\)
    if \(r_{m t, k} \leq \frac{c \cdot \tau_{\text {direct }}}{\sqrt{10}}\) and \(\left|R\left(\xi_{t, k}\right)_{s, r}\right|^{2} \geq \Delta\) then
    \(d\left(r_{m t, k}\right) \leftarrow d\left(\xi_{t, k} \cdot c\right)\)
    otherwise, \(\hat{\tau} \leftarrow\) select delays of candidate echos that satisfy both the SPL and the SE constraits
    for all the delay samples of \(\xi_{t, k}\) and \(\hat{\tau}\) do
    \(\xi_{t, k} \leftarrow \underset{t}{\operatorname{argmin}}\left(\xi_{t, k}-\hat{\tau}\right)\)
    return \(\xi_{t, k}\)
    end for
    \(d\left(r_{m t, k}\right) \leftarrow d\left(\xi_{t, k} \cdot c\right)\)
    end if
```


## 5. LMA-WNLS-Based Pedestrian Self-Positioning

Based on the weighted non-linear least squares (WNLS) approach, the cost function is:

$$
\begin{equation*}
\varepsilon\left(S_{t}\right)=\left(d\left(r_{m t, k}\right)-d\left(r_{e t, k}\right)\right)^{T} * D^{-1} *\left(d\left(r_{m t, k}\right)-d\left(r_{e t, k}\right)\right) \tag{19}
\end{equation*}
$$

where $(\cdot)^{T}$ is the transpose operation, $(\cdot)^{-1}$ is the inverse operation, and $D$ is the noise covariance matrix. $D=\sigma^{2} I_{K-1}$, where $\sigma^{2}$ is the noise covariance and $I$ is the identity matrix. As estimated distances $r_{e t, \mathrm{k}}$ and measured distances $r_{m t, \mathrm{k}}$ are solved by the steps introduced in Sections 3 and 4 , the optimal pedestrian position $\bar{S}_{t}$ is:

$$
\begin{equation*}
\bar{S}_{t}=\operatorname{argmin}_{S_{t}} \varepsilon\left(S_{t}\right) \tag{20}
\end{equation*}
$$

However, the main limitation of the WNLS is that, in order to maintain optimal robustness, its learning rate parameters are usually set to small positives, resulting in a slower convergence rate. Thus, the application of the Levenberg-Marquardt algorithm (LMA) to WNLS could accelerate the convergence while ensuring robustness, and satisfy real-time positioning requirements.

## 6. Experiment

We validated the proposed approach with the data collected from the corridor of the fifth floor of the Jinji Campus Library in GUET, GuiLin, Guangxi Zhuang Autonomous Region, China. The cloister size was $\left[L_{x}, L_{y}\right]=[19,35]$. The four sides of the library corridor are doors, glass windows, and walls; the ceiling is mainly glass with steel stent supports; and the floor is covered with ordinary tile. The whole corridor is a rectangular ring.

The data collection tool used in this experiment was a Huawei Rongyao 7 smartphone installed with a chirp application developed by our team and already authorized by China National Intellectual Property Administration, which was used to emit and store the chirp sound signal. The chirp sample frequency was set as $f_{\mathrm{s}}=44.1 \mathrm{kHz}$, the duration was $T=0.006 \mathrm{~s}$, the lower frequency was $f_{0}=16 \mathrm{kHz}$, the upper frequency was $f_{1}=22 \mathrm{kHz}$, and the emitting interval was 0.3 s . The PDR sample frequency was set as $f_{\mathrm{pdr}}=20 \mathrm{~Hz}$. The empirical energy threshold was set as $\Delta=0.01$.

We had the loudspeaker of the smartphone facing the nearest wall, opened the chirp application, and then walked normally from the starting point (green dot) at $[1.5,9]$ along the corridor to the end point (red dot) at $[1.5,5]$. During data collection, students and staff walked around normally as usual.

### 6.1. Calculation of PDR Information ( $L_{\text {tra }}$ and $\theta$ )

To obtain the adaptive step length $L_{\text {tra }}$, the pedestrian acceleration (denoted $a_{\text {NORM }}$ ) was calculated from the norm of the three-axis accelerometer (denoted $\left\|a_{\kappa}\right\|$ ):

$$
\begin{equation*}
a_{N O R M}=\left\|a_{\kappa}\right\|=\sqrt{\left(a_{x, \kappa}\right)^{2}+\left(a_{y, \kappa}\right)^{2}+\left(a_{z, \kappa}\right)^{2}} \tag{21}
\end{equation*}
$$

where $a_{x, \kappa}, a_{y, k}$ and $a_{z, k}, \kappa=0, \ldots, \mathcal{K}$ are the three-axis accelerometer readings. Then, the sliding window summing technique was used to reduce noise:

$$
\begin{equation*}
S W S(\kappa)=\sum_{t=\kappa-\mathcal{N}+1}^{\kappa} a_{N O R M}(t) \tag{22}
\end{equation*}
$$

where $S W S$ is the sliding window summing, and the window's size was set as $\mathcal{N}=10$. Since $S W S$ is affected by walk motion and gravity, the acceleration differential technique was used to obtain the acceleration differential $a(\kappa)$, as shown in Figure 4:

$$
\begin{equation*}
a(\kappa)=S W S(\kappa+\mathcal{N})-S W S(\kappa) \tag{23}
\end{equation*}
$$



Figure 4. Illustration of $a(\kappa)$, which is the acceleration pattern of a pedestrian in walking states. The zero crossing points, shown in red rectangles, are the detected steps.

Using the acceleration measurements, step detection and step length estimation can be accomplished through the walking frequency $W F$ and acceleration variance $A V$ :

$$
\begin{equation*}
W F=1 /\left(t_{\kappa}-t_{\kappa-1}\right), A V=\left(\sum_{\kappa=1}^{\mathcal{M}}(a(\kappa)-\overline{a(\kappa)})\right) /(\mathcal{M}-1) \tag{24}
\end{equation*}
$$

where $\mathcal{M}$ and $\overline{a(\kappa)}$ are the number of samples and the acceleration mean during a step, respectively. Finally, we obtain $\mathcal{K}=140$ from counting the peaks over zero in $a(\kappa)$ using the find-peaks function. The $L_{\text {tra }}$ plot is shown in Figure 5, and the $\theta$ plot is shown in Figure 6, which was generated using the method described in Section 4.2.


Figure 5. Correlation among $L_{t r a}, \mathcal{K}$, and step frequency. For each step index $\kappa$, the correlation between the $L_{t r a}$ and the step frequency satisfied the statistical theory that the step frequency is larger and the step length is longer.


Figure 6. Heading angular $\theta$ and the corner step index. The heading angular $\theta$ at each $\kappa$ is shown in a blue full line to clarify the $\kappa$ of every room corner. FIR technology is applied. The index values of the corridor corners $\kappa$ are shown in the black numbers.

### 6.2. First-Order Echo Measurements $d\left(r_{m t, k}\right)$

When walking along the corridor, the changing trends of distances from the sound source to the four walls were directly reflected in the values of $r_{m t, k}$, as shown in Figure 7.


Figure 7. Illustration of the trend diagram of $r_{m t, k}$.
Firstly, from step $\kappa=1$ to step $\kappa=40$ (the first corner), the user moved from the south to the north. During this phase, the distance from the east wall and the west wall should remain unchanged, the distance from the south wall should be increasingly larger, and the distance to the north wall should be increasingly smaller. Thus, the trajectory trend when $1 \leq \kappa \leq 40$ was gentle for $k=1$ and $k=3$, increasing for $k=2$, and decreasing for $k=4$.

Next, the user moved from the west to the east; that is, from the first corner $(\kappa=40)$ to the second corner $(\kappa=62)$. During this phase, the distance from the south wall and the north wall should remain unchanged, the distance from the west wall should increase, and the distance to the east wall should decrease. Thus, the trajectory trend when $40 \leq \kappa \leq 62$ is gentle for $k=2$ and $k=4$, increasing for $k=1$, and decreasing for $k=3$.

Similarly, the trends of the distance changes for other sections were the same as the changes in the actual distances.

However, the change parts marked with the black dotted rectangles at every corner point in Figure 7, which should be the smooth transition curves, become sudden sharp declines. After repeating the measurements, we think that the reason for this change is the remaining accumulated errors of the heading angle $\theta$ due to the assumption that the experimenters in this study walked strictly along the dominant directions. In fact, the randomness of a person's walking causes their direction of travel to deviate from the dominant direction, and this error is also eventually reflected in the trajectory of the position tracking.

To further explain the $r_{m t, k}$ extracted from the cross-correlation $R(\tau)_{s, r}$, Figure 8 shows the $r_{m t, k}$ in the $R(\tau)_{s, r}$ when $\kappa=140, k=1$.


Figure 8. $R(\tau)_{s, r}$ when $\kappa=140, k=1$. The $R(\tau)_{s, r}$ wave is generated when the step index of the pedestrian is $\kappa=140$, the impulses marked in red dot lines are the reflections from the west wall.

1. The direct path impulse was found at the peak with maximum value $\left|R(\tau)_{s, r}\right|$, which is marked by D : (X : 1856, Y : 113.5);
2. We subtracted $R(\tau)_{s, s}$ from $R(\tau)_{s, r}$ to eliminate the waveform sidelobe effect and amplify the reverberation parts, as shown in the lower right corner of Figure 8, to find the real first-order echo impulses;
3. Since $\kappa=140>127$, we deduced that the pedestrian has passed the fourth corner and should be on the west side of the corridor, so the peak marked with $\mathrm{M}:(\mathrm{X}: 2344, \mathrm{Y}: 0.02433$ ) generated by the $r_{m t, k}$ at this moment was taken as the first-order reflection from the west wall (i.e., the closest wall); however, $|\mathrm{Y}|^{2} \approx 0.0006 \ll \Delta=0.01$, the measured result did not meet the SE constraint, and so should be updated;
4. With the measured $r_{m t, 3}$ and the constraint of $r_{m t, 1} \approx L_{x}-r_{m t, 3}$, based on the proposed algorithm, the first-order reflection peak related to $r_{m t, 1}$ was updated with the value marked with U : ( $\mathrm{X}: 2286, \mathrm{Y}: 0.1391$ ), which had a smaller distance error than the one before the update, thereby reducing the error of the position; the other first-order reflection peaks were gradually found, and updated.

### 6.3. Self-Positioning Trajectory Comparison

To highlight the advantages of our proposed continuous sound source self-positioning solution, we used two strategies: PDR and our proposed system. The compared results are shown in Figure 9. The following can be seen from the figure: (1) The output of the PDR trajectory (the red line) is continuous and has a similar shape to the reference trajectory (the gray line), but as time increased and the number of pedestrian steps increased, accumulative errors occurred in the accelerometer and gyroscope, resulting in positioning failure. (2) The proposed system output (the blue short line) is closer to the reference trajectory, because it accounts for the acoustic constraints to confirm the required $K=4$ dimension distances between the sound source and its image sources, increasing the accuracy of the positioning result, determined by the starting point to the first corner point, and the trajectory is closer to the reference trajectory.


Figure 9. Trajectory comparison of the pedestrian dead reckoning (PDR) method and our proposed method.

For the same reason as mentioned above, due to the inherent defect of the angle estimation method (an angular cumulative error that cannot be totally eliminated), there were some fluctuations in the corner areas in the tracking trajectory, which is consistent with the change parts marked with the black dotted rectangles in Figure 7, but overall, it was closer to the reference trajectory.

### 6.4. Error Analysis

The errors presented in Figure 8 are illustrated in Figure 10 with the following outcomes:
(1) When $\kappa$ increased, the positioning error increased, as shown in Figure 10a. The error could be as great as 0.5446 m , but the probability was rather low (w.r.t. $1 / 140=0.71 \%$ );
(2) As shown in Figure 10b, the errors of each step were centralized by the histfit function, the probability of error below 0.12 m was $54.46 \%$, and the probability of the error exceeding 0.44 m did not exceed $15.32 \%$;
(3) The box figure (Figure 10c) details the median, maximum, and minimum of the proposed system errors. This result proves that the proposed system is reliable.


Figure 10. Analysis of the proposed system errors. (a) Error values of every step. Here, the step number $\mathcal{K}=140 ;(\mathbf{b})$ error probability distribution histogram produced by the MATLAB histfit function based on the data of (a); and (c) result of the MATLAB box function based on the data of (a), describing the distribution of the collected data and visualizing the normalities and abnormalities of the data.

## 7. Conclusions

We proposed a sensor-rich smartphone-based indoor pedestrian self-positioning system for continuous position acquisition based on image acoustic source impulse. Along with the processing, an acoustic principle-based constraint algorithm was proposed to update the first-order echo measurements generated from the PDR and ITM methods, increasing the reliability of the final positioning results compared to the PDR method. Additionally, the LMA-WNLS model was adopted to reduce the computational complexity of the continuous self-positioning process, thereby increasing time efficiency. Despite this, we noticed some limitations of this system. For example, the used smartphone must have an application that can emit and receive chirp sounds because it is impossible for ordinary smartphones to play chirp sound signals. The arbitrariness of pedestrian motion during walking is limited. If the actual trajectory of walking deviates from the dominant direction, heading angle errors are produced, resulting in positioning error.

Related future work will mainly focus on the data processing of the heading angle and the separation of the close echo arrivals, to further improve the positioning accuracy and fully port this complete system to a smartphone application.

Author Contributions: X.S. designed and performed the experiments and wrote the paper. M.W., H.Q. and L.L. assisted with conceiving of the idea and proofreading of the paper.

Funding: This work was funded by the National Natural Science Foundation of China, grant number No. 61771151, Guangxi Zhuang Autonomous Region Natural Science Foundation, grant number No. 2016GXNSFBA38014, GUET Excellent Graduate Thesis Program, grant number 16YJPYBS02 and Guangxi Experiment Center of Information Science, grant number PT1604.

Acknowledgments: This work was supported by the Ministry of Education Key Laboratory of Cognitive Radio and Information Processing, the Wireless Broadband and Signal Processing Guangxi Key Laboratory.

Conflicts of Interest: The authors declare no conflict of interest.

## References

1. Zhang, R.; Bannoura, A.; Höflinger, F.; Reindl, L.M.; Schindelhauer, C. Indoor localization using a smart phone. In Proceedings of the 2013 IEEE Sensors Applications Symposium Proceedings, Galveston, TX, USA, 19-21 February 2013; pp. 38-42.
2. Liu, J.; Chen, R.; Pei, L.; Guinness, R.; Kuusniemi, H. A Hybrid Smartphone Indoor Positioning Solution for Mobile LBS. Sensors 2012, 12, 17208-17233. [CrossRef] [PubMed]
3. Parviainen, M.; Pertilä, P. Self-localization of dynamic user-worn microphones from observed speech. Appl. Acoust. 2017, 117, 76-85. [CrossRef]
4. Cobos, M.; Perez-Solano, J.J.; Belmonte, Ó.; Ramos, G.; Torres, A.M. Simultaneous Ranging and Self-Positioning in Unsynchronized Wireless Acoustic Sensor Networks. IEEE Trans. Signal Process. 2016, 64, 5993-6004. [CrossRef]
5. Naseri, H.; Koivunen, V. Cooperative Simultaneous Localization and Mapping by Exploiting Multipath Propagation. IEEE Trans. Signal Process. 2016, 65, 200-211. [CrossRef]
6. Bo, C.; Li, X.Y.; Jung, T.; Mao, X. SmartLoc: Sensing landmarks silently for smartphone-based metropolitan localization. Eurasip J. Wirel. Commun. Netw. 2016, 2016, 111. [CrossRef]
7. Lee, H.; Kim, T.H.; Choi, J.W.; Choi, S. Chirp signal-based aerial acoustic communication for smart devices. In Proceedings of the 2015 IEEE Conference on Computer Communications (INFOCOM), Hong Kong, China, 26 April-1 May 2015; pp. 2407-2415.
8. Höflinger, F.; Zhang, R.; Hoppe, J.; Bannoura, A.; Reindl, L.M.; Wendeberg, J.; Bührer, M.; Schindelhauer, C. Acoustic Self-calibrating system for indoor smartphone tracking. In Proceedings of the 2012 International Conference on Indoor Positioning \& Indoor Navigation, Sydney, Australia, 13-15 November 2012; pp. 1-9.
9. Ayllón, D.; Sánchez-Hevia, H.A.; Gil-Pita, R.; Manso, M.U.; Zurera, M.R. Indoor blind localization of smartphones by means of sensor data fusion. In Proceedings of the IEEE Sensors Applications Symposium, Zadar, Croatia, 13-15 April 2015; pp. 1-6.
10. Tarzia, S.P.; Dinda, P.A.; Dick, R.P.; Memik, G. Demo: Indoor localization without infrastructure using the acoustic background spectrum. In Proceedings of the International Conference on Mobile Systems, Applications, and Services, Bethesda, MD, USA, 28 June-1 July 2011; pp. 385-386.
11. Aloui, N.; Raoof, K.; Bouallegue, A.; Letourneur, S.; Zaibi, S. Performance evaluation of an acoustic indoor localization system based on a fingerprinting technique. Eurasip J. Adv. Signal Process. 2014, 2014, 13. [CrossRef]
12. Leonardo, R.; Barandas, M.; Gamboa, H. A Framework for Infrastructure-Free Indoor Localization based on Pervasive Sound Analysis. IEEE Sens. J. 2018, 18, 4136-4144. [CrossRef]
13. Wang, B.; Chen, Q.; Yang, L.T.; Chao, H.C. Indoor smartphone localization via fingerprint crowdsourcing: Challenges and approaches. IEEE Wirel. Commun. 2016, 23, 82-89. [CrossRef]
14. Gutierrez, N.; Belmonte, C.; Hanvey, J.; Espejo, R.; Dong, Z. Indoor localization for mobile devices. In Proceedings of the 11th IEEE International Conference on Networking, Sensing and Control, Miami, FL, USA, 7-9 April 2014; pp. 173-178.
15. Vinyals, O.; Martin, E.; Friedland, G. Multimodal Indoor Localization: An Audio-Wireless-Based Approach. In Proceedings of the IEEE 4th International Conference on Semantic Computing, IEEE Computer Society, Pittsburgh, PA, USA, 22-24 September 2010; pp. 120-125.
16. Bordoy, J.; Wendeberg, J.; Schindelhauer, C.; Reindl, L.M. Single transceiver device-free indoor localization using ultrasound body reflections and walls. In Proceedings of the 2015 International Conference on Indoor Positioning and Indoor Navigation, Banff, AB, Canada, 13-16 October 2015; pp. 1-7.
17. El Badawy, D.; Dokmanić, I.; Vetterli, M. DoA Estimation by One Unsophisticated Sensor. In Proceedings of the 13th International Conference on Latent Variable Analysis and Signal Separation, Grenoble, France, 21-23 February 2017.
18. Parhizkar, R.; Dokmanic, I.; Vetterli, M. Single-channel indoor microphone localization. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Florence, Italy, 4-9 May 2014; pp. 1434-1438.
19. Dokmanić, I.; Daudet, L.; Vetterli, M. From acoustic room reconstruction to slam. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20-25 March 2016; pp. 6345-6349.
20. Kreković, M.; Dokmanić, I.; Vetterli, M. EchoSLAM: Simultaneous localization and mapping with acoustic echoes. In Proceedings of the IEEE International Conference on Acoustics, Speech and Signal Processing, Shanghai, China, 20-25 March 2016; pp. 11-15.
21. Qian, J.; Pei, L.; Ma, J.; Ying, R.; Liu, P. Vector graph assisted pedestrian dead reckoning using an unconstrained smartphone. Sensors 2015, 15, 5032-5057. [CrossRef] [PubMed]
22. Tian, Q.; Salcic, Z.; Wang, K.I.K.; Pan, Y. A Multi-Mode Dead Reckoning System for Pedestrian Tracking Using Smartphones. IEEE Sens. J. 2016, 16, 2079-2093. [CrossRef]
23. Yang, H.; Zhang, R.; Bordoy, J.; Höflinger, F.; Li, W.; Schindelhauer, C.; Reindl, L. Smartphone-Based Indoor Localization System Using Inertial Sensor and Acoustic Transmitter/Receiver. IEEE Sens. J. 2016, 16, 8051-8061. [CrossRef]
24. Foxlin, E. Pedestrian Tracking with shoe-mounted inertial sensors. IEEE Comput. Graph. Appl. 2005, 25, 38-46. [CrossRef] [PubMed]
25. Ruiz, A.R.J.; Granja, F.S.; Honorato, J.C.P.; Rosas, J.I.G. Accurate Pedestrian Indoor Navigation by Tightly Coupling Foot-Mounted IMU and RFID Measurements. IEEE Trans. Instrum. Meas. 2011, 61, 178-189. [CrossRef]
26. Xaver, F.; Matz, G.; Gerstoft, P.; Mecklenbräuker, C. Localization of acoustic sources using a decentralized particle filter. Eurasip J. Wirel. Commun. Netw. 2011, 2011, 94 . [CrossRef]
27. Ruotsalainen, L.; Kirkko-Jaakkola, M.; Rantanen, J.; Mäkelä, M. Error Modelling for Multi-Sensor Measurements in Infrastructure-Free Indoor Navigation. Sensors 2018, 18, 590. [CrossRef] [PubMed]
28. Borish, J. Extension of the image model to arbitrary polyhedral. J. Acoust. Soc. Am. 1984, 75, 1827-1836. [CrossRef]
29. Fu, Z.H.; Li, J.W. GPU-based image method for room impulse response calculation. Multimedia Tools Appl. 2016, 75, 5205-5221. [CrossRef]
30. Shin, S.H.; Chan, G.P. Adaptive step length estimation algorithm using optimal parameters and movement status awareness. Med. Eng. Phys. 2011, 33, 1064-1071. [CrossRef] [PubMed]
31. Deng, Z.; Cao, Y.; Wang, P.; Wang, B. An Improved Heuristic Drift Elimination Method for Indoor Pedestrian Positioning. Sensors 2018, 18, 1874. [CrossRef] [PubMed]
32. Mensing, C.; Plass, S. Positioning Algorithms for Cellular Networks Using TDOA. In Proceedings of the IEEE International Conference on Acoustics Speech and Signal Processing Proceedings, Toulouse, France, 14-19 May 2006.
33. Rife, D.D. Transfer-function measurement with maximum-length sequences. J. Audio Eng. Soc. 1989, 37, 102-113.
34. Knapp, C.; Carter, G. The generalized correlation method for estimation of time delay. IEEE Trans. Acoust. Speech Signal Process. 2003, 24, 320-327. [CrossRef]
35. Dibiase, J.H.; Silverman, H.F.; Brandstein, M.S. Robust Localization in Reverberant Rooms. In Microphone Arrays; Springe: Berlin/Heidelberg, Germany, 2001.
36. Chrisler, V.L. Sound Absorption Coefficients. J. Acoust. Soc. Am. 1934, 6, 115. [CrossRef]
© 2018 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (http:// creativecommons.org/licenses/by/4.0/).
