

Two-stage Localisation Scheme Using a Small-scale Linear Microphone Array for Indoor Environments

Wuyang Jiang¹, Ling Pei¹, Changqing Xu¹, Liang Chen² and Wenxian Yu³

¹(Shanghai Key Laboratory of Navigation and Location Based Services, Shanghai Jiao Tong University, China)

²(Finnish Geodetic Institute, Masala, Finland)

³(Shanghai Key Laboratory of Intelligent Sensing and Recognition, Shanghai Jiao Tong University, China)

(E-mail: ling.pei@sjtu.edu.cn)

The small-scale linear microphone arrays that are widely found in smartphones could be used to locate a sound source for indoor environments. After the Time Differences Of Arrival (TDOAs) in microphone pairs are estimated, a TDOA-based hybrid localisation scheme is proposed for a small-scale linear microphone array. The scheme contains two stages: the initialisation stage using the Levenberg-Marquardt (LM) algorithm, and the refining solution stage using the Weighted Least-Square (WLS) algorithm or the Multi-Dimensional Scaling-based (MDS) algorithm. Simulations and field tests show that the proposed indoor localisation scheme outperforms the existing schemes, and it can achieve an average error of 0.32 metres in an 8 m by 5 m area.

KEYWORDS

1. Localisation. 2. TDOA. 3. Linear. 4. Small-scale

Submitted: 30 September 2014. Accepted: 20 January 2015. First published online: 23 February 2015.

1. INTRODUCTION. Localisation technologies and location-based applications have been extensively investigated in recent years. Global Navigation Satellites Systems (GNSS) have been widely adopted for positioning and navigation in outdoor scenarios. Indoor environments, on the other hand, bring about more location-based services opportunities as well as more challenges for localisation because of less availability of navigation satellites and complicated built structures hindering signal propagation.

Other than GNSS, short-ranged signals such as WiFi, Bluetooth (Bahl and Padmanabhan, 2000; Chen et al., 2013; Gomes and Sarmiento, 2009; Pei et al., 2010b; Pei et al., 2010c; Priyantha et al., 2000), Radio-Frequency Identification (RFID) (Ni et al., 2004), Ultra Wideband (UWB) (Pahlavan et al., 2006), long-ranged signals such as Global System for Mobile communications (GSM) (Syrjärinne, 2001), Digital Television (DTV) (Chen et al., 2012), and magnetic fields

(Storms et al., 2010) have been utilised in indoor localisation for schemes such as proximity, fingerprinting, or triangulation. Microelectromechanical Systems (MEMS) motion sensors offer the opportunity of continuous relative navigation when the localisation infrastructure is unavailable (Chen et al., 2010; Chen et al., 2011a; Chen et al., 2011b; Foxlin, 2005; Indoor Atlas Ltd, 2011; Mathews et al., 2011; Pei et al., 2010a; Pei et al., 2011; Pei et al., 2012; Pei et al., 2013). The low cost camera is also a potential localisation sensor (Ruotsalainen et al., 2011). Furthermore, hybrid solutions (Kuusniemi et al., 2012; Liu et al., 2010; Liu et al., 2012a; Liu et al., 2012b) are adopted to improve the availability and reliability of localisation.

Today, the small-scale linear microphone arrays widely installed on smartphones could be used for localisation. Although the microphones involved in a small-scale array are physically located in one single device of a portable size that is easy to carry, their small scale poses major challenges to the existing localisation technologies. While large-scale microphone arrays have been widely applied for localisation (Filonenko et al., 2010; Filonenko et al., 2012; Hoflinger et al., 2012; Holm, 2012; Janson et al., 2010; Pertila et al., 2012), small-scale linear microphones are mainly used for discriminating sounds based on direction, locating sound sources in terms of direction, and enhancing the microphone performance by reducing noises. For instance, microphone arrays have been used for scream and gunshot detection and localisation (Valenzise et al., 2007), for monitoring vehicles, e.g. airplane, vessel, auto, etc., for noise diagnosing (Michel and Barsikow, 2003; Piet et al., 2002; Schulte-Werning et al., 2003), noise reduction, speech recognition, teleconference, speaker tracking (Khalil et al., 1994), and for game consoles, etc. However, the possibility of localisation using a small-scale array is much less explored in the literature.

In this paper, we propose an indoor localisation scheme, which relies on a small-scale linear microphone array to locate a sound source based on Time Difference Of Arrival (TDOA). The proposed localisation scheme is a hybrid solution on the basis of three existing TDOA localisation methods, which are the Weighted Least-Square (WLS) method, the Multi-Dimensional Scaling-based (MDS) method and the Levenberg-Marquardt (LM) method. For the existing localisation methods, we present two findings by simulations and field tests: 1) the WLS method and the MDS method perform well only for sources located close to the perpendicular bisector of the microphone array, while the MDS method is slightly better than the WLS method, and 2) the LM method can give an initial estimate, which indicates the location for a near-field source, or the direction for a far-field source. Considering these new findings, we design a hybrid localisation scheme consisting of two stages: 1) find the initial estimate of the source using the LM method, 2) rotate the microphone array so that the source can be roughly relocated on the perpendicular bisector of the microphone array, then find the refining estimate of the location of the source using the MDS method. By this two-stage procedure, the scheme has the advantages of the existing methods and efficiently expands the feasible area of the localisation scheme using a small-scale linear microphone array. Simulations and field tests results show that the proposed scheme can achieve an average error of 0.32 m in an 8 m by 5 m area.

The rest of the paper is organised as follows. The fundamentals of a small-scale linear microphone array are described in Section 2. Section 3 introduces the TDOA approach adopted in this paper. In Section 4, we present the proposed localisation algorithms. Our experiments and results are discussed in Section 5. Finally, some concluding remarks are drawn in Section 6.

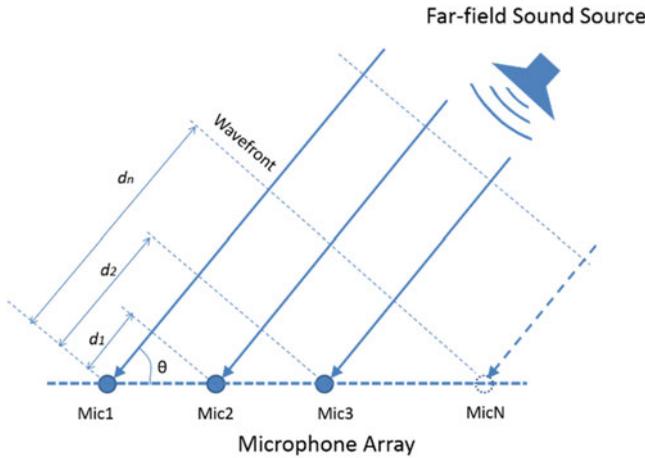


Figure 1. Far-field model.

2. SMALL-SCALE LINEAR MICROPHONE ARRAY

2.1. *Speed of Sound.* Sound is a compression wave that travels around 342 metres per second in dry air with temperature at 20 °C. The speed of sound in conditions of various temperatures can be calculated from Hoflinger et al. (2012):

$$C = 331.3 \times \sqrt{1 + T/273.15}, \tag{1}$$

where C is the speed of sound in m/s, and T is the temperature in °C. In contrast with electromagnetic waves, the propagation speed of sound is significantly slower, which relaxes the requirement of timing in a sound-based localisation system.

2.2. *Far-field and Near-field.* For a linear array, a sound may be considered to come from a far-field source as shown in Figure 1 if:

$$|r| > \frac{2L^2f}{C}, \tag{2}$$

where r is the distance from a sound source to the centre of the microphone array, L is the dimension of the linear microphone array, f is the frequency of the sound, and C is the speed of sound (Steinberg, 1976). The near-field sound source is shown in Figure 2.

3. TIME DELAY ESTIMATION

3.1. *TDOA.* The range from the sound source to the i th microphone of the microphone array can be calculated as:

$$r_i = Ct_i, \tag{3}$$

where t_i is the travel time of sound from the sound source to the i th microphone. A TDOA solution does not require the synchronisation of the clock of the sound source and that of the microphone array, which reduces the system complexity.

Suppose that the three-dimensional coordinates of the sound source and the microphones are denoted as $\mathbf{s} = [s_x, s_y, s_z]^T \in \mathbf{R}^3$ and $\mathbf{m}_k = [m_{x_k}, m_{y_k}, m_{z_k}]^T \in \mathbf{R}^3$,

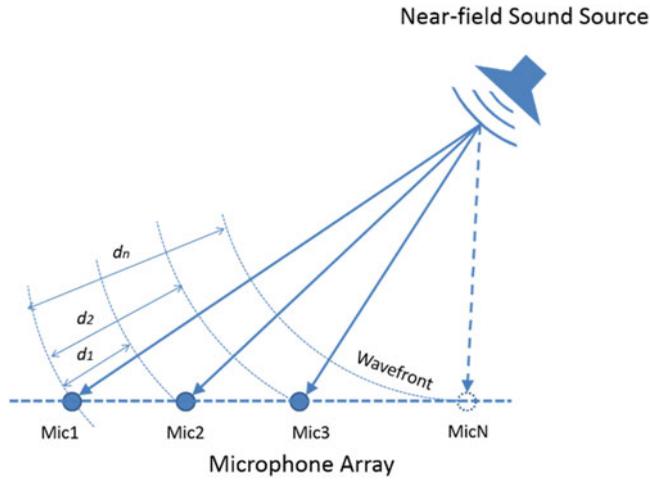


Figure 2. Near-field model.

respectively. Then the ideal TDOA associated with the i th microphone pair ($\mathbf{m}_{i1}, \mathbf{m}_{i2}$) is computed as (Brandstein, 1995):

$$\Delta t_i = \frac{\Delta r_i}{C} = \frac{\|\mathbf{s} - \mathbf{m}_{i1}\| - \|\mathbf{s} - \mathbf{m}_{i2}\|}{C}, \tag{4}$$

where $\|\cdot\|$ denotes the Euclidian distance between two points. Δr_i is the range difference between distance $\|\mathbf{s} - \mathbf{m}_{i1}\|$ and $\|\mathbf{s} - \mathbf{m}_{i2}\|$. Equation (4) indicates that we can solve the unknown location of a sound source with TDOA measurements of at least three independent microphone pairs.

3.2. *Time Delay Estimation.* Considering that in practice the TDOA estimate is a non-ideal measurement around the ideal TDOA, we formulate the Time Delay Estimation (TDE) from Equation (4) as:

$$\Delta \hat{t}_i = \frac{\|\mathbf{s} - \mathbf{m}_{i1}\| - \|\mathbf{s} - \mathbf{m}_{i2}\|}{C} + \eta_i, \tag{5}$$

where η_i is an additive error assumed to be independently distributed for different microphone pairs.

The microphone signals at the i th and the j th microphones can be respectively modelled as:

$$\begin{cases} x_i(t) = s(t) + n_i(t) \\ x_j(t) = s(t + \Delta t_{ij}) + n_j(y) \end{cases}, \tag{6}$$

where $s(t)$ is the source signal, $n_i(t)$ and $n_j(t)$ represent the additive noises respectively at the i th and the j th microphones, Δt_{ij} denotes the time delay between two received signals. The aim of TDE is to estimate Δt_{ij} that is proportional to the range difference (Brandstein and Ward, 2001; Carter, 1987; Knapp and Carter, 1976). The time domain

cross-correlation method obtains the delay estimate between two microphones as the time lag that maximizes the cross-correlation between the received signals. The frequency domain algorithms based on the Generalised Cross-Correlation (GCC) method implement a frequency domain cross-correlation by introducing a weighting function to eliminate noise disturbances:

$$R_{x_1x_2}(\tau) \equiv \frac{1}{2\pi} \int_{-\infty}^{\infty} W(\omega)X_1(\omega)X_2^*(\omega)e^{j\omega\tau} d\omega, \tag{7}$$

where $X_1(\omega)$, $X_2(\omega)$ are the Fourier transforms of two microphone signals. $X_2^*(\omega)$ is the conjugate function of $X_2(\omega)$. The Maximum Likelihood (ML) approach and the Phase Transform (PHAT) weighting functions are two widely used functions. The ML approach performs well in moderately noisy and non-reverberant environments (DiBiase et al., 2001; Ianniello, 1982). In environments with high reverberations, the PHAT weighting function (Knapp and Carter, 1976), shown in Equation (8), has more robust performance:

$$W_{PHAT}(\omega) = \frac{1}{|X_i(\omega)X_j^*(\omega)|}. \tag{8}$$

4. LOCALISATION ALGORITHMS. In this section, firstly we address the problem of sound localisation based on the TDOA scheme. Secondly we adapt three basic localisation methods to solve the localisation problem for a linear array, namely the WLS method (Chan and Ho, 1994), the MDS method (Wei et al., 2008; Wei et al., 2010), and the LM method (Lourakis, 2005; Moré, 1978; Nocedal and Wright, 1999). Finally, we propose a hybrid localisation scheme to solve the localisation problem for a linear array. Our work is derived in a 2-D plane scenario, where the axis parallel to the linear array is set as the X-axis, and the perpendicular bisector of the linear array as the Y-axis. Extension to 3-D scenarios remains for future studies.

4.1. *Problem Statement.* The localisation problem can be formulated by over-determined non-linear equations. Given a specific speed of sound, the TDOA transforms to the range difference of arrival. Selecting the first microphone as reference, the range differences of all the possible microphone pairs with respect to the first microphone can be defined as:

$$\Delta r_{k,1} = r_k - r_1 \equiv C\Delta t_{k,1}, k = 2, 3, \dots, M. \tag{9}$$

Taking into account the measurement error of time differences of arrival and from Equation (5), we rewrite Equation (9) as:

$$\Delta r_{k,1} \approx C\Delta \hat{t}_{k,1}, k = 2, 3, \dots, M. \tag{10}$$

For convenience, we define $\Delta r_{1,1} = 0$.

On the other hand, r_k can be expressed as:

$$r_k = \sqrt{(m_{x_k} - s_x)^2 + (m_{y_k} - s_y)^2}, k = 2, 3, \dots, M. \tag{11}$$

Then, the unknown sound source location $\mathbf{s} = [s_x, s_y]^T$ should be the best approximate solution to the following set of non-linear equations:

$$\sqrt{(m_{x_k} - s_x)^2 + (m_{y_k} - s_y)^2} - \sqrt{(m_{x_1} - s_x)^2 + (m_{y_1} - s_y)^2} = C\Delta t_{k,1}, \tag{12}$$

$$k = 2, 3, \dots, M.$$

4.2. *Basic Localisation Algorithms.* We consider three basic methods, i.e., the WLS method, the MDS method and the LM method to solve the localisation problem. The detailed algorithms of the three methods are shown as Algorithms 1, 2, and 3, respectively. While Algorithm 1 is a direct application of existing literature, Algorithms 2 and 3 are particularly modified from existing literature to adapt to the usage of the linear array. Also, Algorithms 1 and 2 are closed-form solutions, and Algorithm 3 is an iterative solution. Some remarks are shown at the end of the algorithms. The simulation results and further discussion are presented in Section 5.

Algorithm 1. The WLS method (application of Chan and Ho (1994))

1: $\rho_k^2 = m_{x_k}^2 + m_{y_k}^2, k = 1, 2, \dots, M.$

2: $x_{k,1} = m_{x_k} - m_{x_1}, k = 2, \dots, M.$

3: $\mathbf{Q} = \begin{bmatrix} 1 & 0.5 & \dots & 0.5 \\ 0.5 & 1 & \dots & 0.5 \\ \vdots & \vdots & \ddots & \vdots \\ 0.5 & 0.5 & \dots & 1 \end{bmatrix}_{(M-1) \times (M-1)}.$

4: $\mathbf{h} = \frac{1}{2} \begin{bmatrix} \Delta r_{2,1}^2 - \rho_2^2 + \rho_1^2 \\ \Delta r_{3,1}^2 - \rho_3^2 + \rho_1^2 \\ \vdots \\ \Delta r_{M,1}^2 - \rho_M^2 + \rho_1^2 \end{bmatrix}.$

5: $\mathbf{G} = - \begin{bmatrix} x_{2,1} & \Delta r_{2,1} \\ x_{3,1} & \Delta r_{3,1} \\ \vdots & \vdots \\ x_{M,1} & \Delta r_{M,1} \end{bmatrix}.$

6: $\mathbf{x}_s^{(1)} = [s_x^{(1)}, r_1^{(1)}]^T = (\mathbf{G}^T \mathbf{Q}^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{Q}^{-1} \mathbf{h}.$

7: $s_y^{(1)} = \sqrt{(r_1^{(1)})^2 - (s_x^{(1)})^2}.$

$$8: \mathbf{B}^{(1)} = \text{diag} \left\{ \sqrt{\left(m_{x_2} - s_x^{(1)}\right)^2 + \left(m_{y_2} - s_y^{(1)}\right)^2}, \right. \\ \left. \sqrt{\left(m_{x_3} - s_x^{(1)}\right)^2 + \left(m_{y_3} - s_y^{(1)}\right)^2}, \dots, \sqrt{\left(m_{x_M} - s_x^{(1)}\right)^2 + \left(m_{y_M} - s_y^{(1)}\right)^2} \right\}.$$

$$9: \Sigma = \mathbf{C}^2 \mathbf{B}^{(1)} \mathbf{Q} \mathbf{B}^{(1)}.$$

$$10: \mathbf{x}_s^{(2)} = \left[s_x^{(2)}, r_1^{(2)} \right]^T = \left(\mathbf{G}^T \Sigma^{-1} \mathbf{G} \right)^{-1} \mathbf{G}^T \Sigma^{-1} \mathbf{h}.$$

$$11: s_y^{(2)} = \sqrt{\left(r_1^{(2)}\right)^2 - \left(s_x^{(2)}\right)^2}.$$

12: The result $\left[s_x, s_y \right]^T = \left[s_x^{(2)}, s_y^{(2)} \right]^T$ is the best approximate solution in this method.

Algorithm 2. The MDS method (modified from Wei et al. (2008))

1: $\mathbf{z}_k = \left[m_{x_k}, m_{y_k}, i \cdot \Delta r_{k,1} \right]^T, k = 1, 2, \dots, M, i^2 = 1.$

2: Calculate the $(M + 1) \times (M + 1)$ matrix \mathbf{D} , whose $(m + 1, n + 1)$ th entry is:

$$\begin{aligned} [\mathbf{D}]_{m+1,n+1} &= (\mathbf{z}_m - \mathbf{z}_n)^T \cdot (\mathbf{z}_m - \mathbf{z}_n) \\ &= (m_{x_m} - m_{x_n})^2 + (m_{y_m} - m_{y_n})^2 - (\Delta r_{m,1} - \Delta r_{n,1})^2, m, n = 1, \dots, M \end{aligned}$$

The entries in the first column and the first row of \mathbf{D} are all set to zero.

3: $\mathbf{J}_{M+1} = \mathbf{I}_{M+1} - \frac{1}{M+1} \mathbf{1}_{M+1} \mathbf{1}_{M+1}^T$, where $\mathbf{1}_{M+1}$ is a $(M + 1)$ dimensional column vector with all elements 1.

$$4: \mathbf{B} = -\frac{1}{2} \mathbf{J}_{M+1} \mathbf{D} \mathbf{J}_{M+1}.$$

5: Find the eigenvalue decomposition of \mathbf{B} , $\mathbf{B} = \mathbf{U} \Lambda \mathbf{U}^T$, where $\Lambda = \text{diag} \{ \lambda_1, \dots, \lambda_{M+1} \}$, $\lambda_1 \geq \dots \geq \lambda_{M+1} \geq 0$, $\mathbf{U} = [\mathbf{u}_1, \dots, \mathbf{u}_{M+1}]$, $\mathbf{U} \mathbf{U}^T = \mathbf{I}$.

6: $\mathbf{U}_n = [\mathbf{u}_4, \dots, \mathbf{u}_{m+1}]$.

$$7: \mathbf{P} = \begin{bmatrix} -\frac{1}{M+1} \sum_{k=1}^M m_{x_k} & -\frac{1}{M+1} \sum_{k=1}^M m_{y_k} & -\frac{1}{M+1} \sum_{k=1}^M \Delta r_{k,1} \\ m_{x_1} - \frac{1}{M+1} \sum_{k=1}^M m_{x_k} & m_{y_1} - \frac{1}{M+1} \sum_{k=1}^M m_{y_k} & \Delta r_{1,1} - \frac{1}{M+1} \sum_{k=1}^M \Delta r_{k,1} \\ \vdots & \vdots & \vdots \\ m_{y_m} - \frac{1}{M+1} \sum_{k=1}^M m_{x_k} & m_{y_m} - \frac{1}{M+1} \sum_{k=1}^M m_{y_k} & \Delta r_{M,1} - \frac{1}{M+1} \sum_{k=1}^M \Delta r_{k,1} \end{bmatrix}.$$

8: $\mathbf{m} = [1/(M + 1) - 1, 1/(M + 1), \dots, 1/(M + 1)]^T$ with $(M + 1) \times 1$ dimensions.

$$9: \mathbf{z} = [s_x, \tilde{s}_y, -r_1]^T = \frac{\mathbf{P}^T \mathbf{U}_n \mathbf{U}_n^T \mathbf{m}}{\mathbf{m}^T \mathbf{U}_n \mathbf{U}_n^T \mathbf{m}}.$$

$$10: s_y = m_{y_1} + \sqrt{r_1^2 - s_x^2}.$$

11: The result $[s_x, s_y]^T$ is the best approximate solution in this method.

Remarks:

The key equation which makes Algorithm 2 valid is $\mathbf{P}^T \mathbf{U}_n = \mathbf{z} \mathbf{m}^T \mathbf{U}_n$, where $\mathbf{z} = [s_x, s_y, -r_1]^T$. Since we set Y-axis as the perpendicular bisector of the linear array, all the $m_{y_k} (k = 1, 2, \dots, M)$ are the same. It can be observed that the second column of \mathbf{P} is proportional to \mathbf{m} , leading to that the second row of $\mathbf{P}^T \mathbf{U}_n$ and $\mathbf{z} \mathbf{m}^T \mathbf{U}_n$ are linearly dependent. So the second element of \mathbf{z} is forced to be a constant, which equals to m_{y_1} . That is to say, s_y cannot be directly solved from \mathbf{z} . Hence, Step 9 of Algorithm 2 only gives a tentative solution for s_y , i.e., \tilde{s}_y . The final solution for s_y is obtained in Step 10 of Algorithm 2.

Algorithm 3. The LM method (modified from Nocedal and Wright (1999))

1: Set $\bar{\Delta} > 0, \Delta_0 \in (0, \bar{\Delta}), \eta \in [0, 1/4)$.

2: Initialize $\mathbf{x}_s^{(0)} = [s_x^{(0)}, s_y^{(0)}]^T$.

3: for $l = 0, 1, 2, \dots$

4: Calculate $\boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)})$, $\mathbf{J}(\mathbf{x}_s^{(l)})$ from Equations (13–17):

$$\varepsilon_k(\mathbf{x}_s^{(l)}) = \left(\sqrt{(s_x^{(l)} - m_{x_k})^2 + (s_y^{(l)} - m_{y_k})^2} - \sqrt{(s_x^{(l)} - m_{x_1})^2 + (s_y^{(l)} - m_{y_1})^2} \right) - \Delta r_{k,1},$$

$$k = 2, \dots, M \tag{13}$$

$$\boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)}) = [\varepsilon_2(\mathbf{x}_s^{(l)}), \varepsilon_3(\mathbf{x}_s^{(l)}), \dots, \varepsilon_M(\mathbf{x}_s^{(l)})]^T, \tag{14}$$

$$\frac{\partial \varepsilon_k(\mathbf{x}_s^{(l)})}{\partial s_x} = \frac{s_x^{(l)} - m_{x_k}}{\sqrt{(s_x^{(l)} - m_{x_k})^2 + (s_y^{(l)} - m_{y_k})^2}} - \frac{s_x^{(l)} - m_{x_1}}{\sqrt{(s_x^{(l)} - m_{x_1})^2 + (s_y^{(l)} - m_{y_1})^2}},$$

$$k = 2, \dots, M \tag{15}$$

$$\frac{\partial \varepsilon_k(\mathbf{x}_s^{(l)})}{\partial s_y} = \frac{s_y^{(l)} - m_{y_k}}{\sqrt{(s_x^{(l)} - m_{x_k})^2 + (s_y^{(l)} - m_{y_k})^2}} - \frac{s_y^{(l)} - m_{y_1}}{\sqrt{(s_x^{(l)} - m_{x_1})^2 + (s_y^{(l)} - m_{y_1})^2}},$$

$$k = 2, \dots, M \tag{16}$$

$$\mathbf{J}(\mathbf{x}_s^{(l)}) = \begin{bmatrix} \frac{\partial \varepsilon_2(\mathbf{x}_s^{(l)})}{\partial s_x} & \frac{\partial \varepsilon_2(\mathbf{x}_s^{(l)})}{\partial s_y} \\ \vdots & \vdots \\ \frac{\partial \varepsilon_M(\mathbf{x}_s^{(l)})}{\partial s_x} & \frac{\partial \varepsilon_M(\mathbf{x}_s^{(l)})}{\partial s_y} \end{bmatrix}. \tag{17}$$

5: Let the diagonal matrix \mathbf{D} with non-negative entrances satisfy:

$$[\mathbf{D}^2]_{n,n} = \left[\mathbf{J}(\mathbf{x}_s^{(l)})^T \cdot \mathbf{J}(\mathbf{x}_s^{(l)}) \right]_{n,n}, \quad n = 1, 2.$$

6: Obtain \mathbf{p}_l that (approximately) satisfies:

$$\min_{\mathbf{p} \in \mathbb{R}^2} m_l(\mathbf{p}) = f(\mathbf{x}_s^{(l)}) + \left(\mathbf{J}(\mathbf{x}_s^{(l)})^T \cdot \boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)}) \right)^T \cdot \mathbf{p} + \frac{1}{2} \mathbf{p}^T \cdot \left(\mathbf{J}(\mathbf{x}_s^{(l)})^T \cdot \mathbf{J}(\mathbf{x}_s^{(l)}) \right) \cdot \mathbf{p}, \quad \text{s.t. } \|\mathbf{D}\mathbf{p}\| \leq \Delta_l.$$

7: Calculate $f(\mathbf{x}_s^{(l)})$, $f(\mathbf{x}_s^{(l)} + \mathbf{p}_l)$ from Equations (13) and (18):

$$f(\mathbf{x}_s) = \frac{1}{2} \sum_{k=2}^M \varepsilon_k^2(\mathbf{x}_s). \tag{18}$$

$$8: \rho_l = \frac{f(\mathbf{x}_s^{(l)}) - f(\mathbf{x}_s^{(l)} + \mathbf{p}_l)}{m_l(\mathbf{0}) - m_l(\mathbf{p}_l)} = \frac{f(\mathbf{x}_s^{(l)}) - f(\mathbf{x}_s^{(l)} + \mathbf{p}_l)}{-\left(\mathbf{J}(\mathbf{x}_s^{(l)})^T \boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)})\right)^T \mathbf{p}_l - \frac{1}{2} \mathbf{p}_l^T \mathbf{J}(\mathbf{x}_s^{(l)}) \mathbf{J}(\mathbf{x}_s^{(l)}) \mathbf{p}_l}.$$

9: If $\rho_l < 1/4$, then $\Delta_{l+1} = 1/4 \cdot \|\mathbf{p}_l\|$.

10: If $\rho_l > 3/4$ and $\|\mathbf{p}_l\| < \Delta_l$, then $\Delta_{l+1} = \Delta_l$.

11: If $1/4 \leq \rho_l \leq 3/4$, then $\Delta_{l+1} = \Delta_l$.

12: If $\rho_l > 3/4$ and $\|\mathbf{p}_l\| = \Delta_l$, then $\Delta_{l+1} = \min(2\Delta_l, \bar{\Delta})$.

13: If $\rho_l > \eta$, then $\mathbf{x}_s^{(l+1)} = \mathbf{x}_s^{(l)} + \mathbf{p}_l$.

14: If $\rho_l \leq \eta$, then $\mathbf{x}_s^{(l+1)} = \mathbf{x}_s^{(l)}$.

15: If one of the following conditions is satisfied, then stop iterating, and let $\hat{\mathbf{x}}_s = \mathbf{x}_s^{(l)}$.

1) $\left\| \mathbf{J}(\mathbf{x}_s^{(l)})^T \cdot \boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)}) \right\| < \text{th}_1$, where th_1 is a preset threshold.

2) $\|\mathbf{p}_l\| < \text{th}_2$, where th_2 is a preset threshold.

3) $f(\mathbf{x}_s^{(l)}) < \text{th}_3$, where th_3 is a preset threshold.

4) The maximum number of iterations is reached.

16: **end (for)**

Remarks:

1) The detail of Step 6 of Algorithm 3 is shown in the Appendix.

- 2) Limited by the objective function, the result of the LM method heavily depends on the initial point.
- 3) The LM method works in both far-field and near-field conditions. If the sound source is a far-field source, the LM method can give an estimate of the direction of the source. If the sound source is a near-field source, the LM method can give an estimate of the location of the source. In contrast, the direction-finding algorithms, such as Multiple Signal Classification (MUSIC), work only in the far-field condition.

4.3. *The Proposed Hybrid Localisation Algorithm.* We propose a hybrid localisation scheme on the basis of the discussed three basic localisation methods.

Algorithm 4. The hybrid scheme

- 1: Use Algorithm 3 (LM) to find an initial estimate of the source location, denoted as $\mathbf{x}_s^{(1)} = [s_x^{(1)}, s_y^{(1)}]^T$.
- 2: Calculate the angle $\theta = \text{angle}(\mathbf{x}_s^{(1)})$. θ refers to the angle from the Y-axis to the line connecting the origin of the coordinate system and the initial estimate of the source location.
- 3: Rotate the microphone array around the origin point by the angle of θ , so that the source can be roughly relocated on the perpendicular bisector of the microphone array.
- 4: Use Algorithm 1 (WLS) or Algorithm 2 (MDS) to find a refining estimate of the source location in the new coordinate system, denoted as $\mathbf{x}_s^{(2)} = [s_x^{(2)}, s_y^{(2)}]^T$.
- 5: Use the standard coordinate transform to obtain the coordinates of the refining estimate of the sound source in the original coordinate system as:

$$\mathbf{x}_s^{(3)} = [s_x^{(3)}, s_y^{(3)}]^T = \begin{bmatrix} \cos \theta & -\sin \theta \\ \sin \theta & \cos \theta \end{bmatrix} \mathbf{x}_s^{(2)}.$$

Remarks:

- 1) Algorithm 4 contains two stages: the initialisation stage (Steps 1 and 2) to estimate the direction of the source, and the refining solution stage (Steps 3, 4, and 5) to estimate the location of the source.
- 2) The purpose of Step 3 of Algorithm 4 is to adjust the posture of the microphone array so that the source is relocated close to the perpendicular bisector of the microphone array, i.e., close to the Y-axis in the new coordinate system. It will be illustrated in Section 5 that only when the source is located close to the Y-axis can it be localised accurately by Algorithms 1 and 2. It is for this reason that Step 3 of Algorithm 4 is needed.
- 3) Step 3 of Algorithm 4 is feasible in practice because the microphone array is a small-scale device whose posture is easy to control.

5. EXPERIMENTAL RESULTS. In this section, via simulations and field tests, we evaluate the performance of the basic localisation methods and the proposed hybrid localisation scheme discussed in Section 4. As an example of a small-scale linear microphone array, a Kinect device (Webb and Ashley, 2012) with four

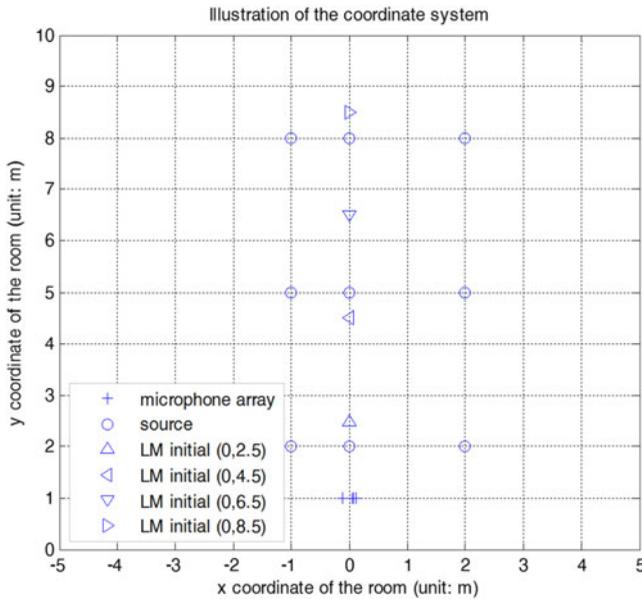


Figure 3. Illustration of the positions of the microphone array, the sources, and the initial points for the LM method.

microphones arranged in a line is used in our field tests. The width of the microphone array is measured to be 0.226 m. Simulations are also conducted with the same microphone deployment adopted in our field test.

5.1. *Simulation Results and Discussions.* We consider a four-element microphone array in a 10 m \times 10 m area. Also, we assume that the sound source and microphones are located at the same height so that the 3-D problem is simplified to a 2-D one. We place the small-scale microphone array on the line of $y = 1$, as shown in Figure 3. To be more specific, the coordinates of the four microphones in the linear array are $(-0.113, 1)$, $(0.036, 1)$, $(0.072, 1)$, $(0.113, 1)$, respectively.

Firstly, we evaluate the performance of the three basic localisation methods described in Algorithms 1, 2, and 3. To illustrate that the localisation performance varies with source locations, we select 9 typical positions, the coordinates of which are $(0, 2)$, $(-1, 2)$, $(2, 2)$, $(0, 5)$, $(-1, 5)$, $(2, 5)$, $(0, 8)$, $(-1, 8)$, $(2, 8)$ as shown in Figure 3. Since the LM method needs to initialise an initial point of searching, we set four different initial points located at $(0, 2.5)$, $(0, 4.5)$, $(0, 6.5)$, $(0, 8.5)$.

In our simulations, we suppose that the estimate error η_i follows an independent and identically distributed (i.i.d.) zero-mean Gaussian distribution. For each algorithm and each sound source position, 100 independent experiments are conducted. In the WLS method and the LM method, the average of the outputs from the 100 experiments is set as the final localisation result. In the MDS method, on the other hand, the median of the outputs from the 100 experiments is set as the final localisation result, so that some experiments with huge estimate errors can be excluded from the performance evaluation. Such huge estimate errors are usually caused by a nearly zero denominator in Step 9 of Algorithm 2. Suppose that the standard deviation (STD) of the assumed Gaussian distribution is $2.924 \mu\text{s}$, which corresponds to a

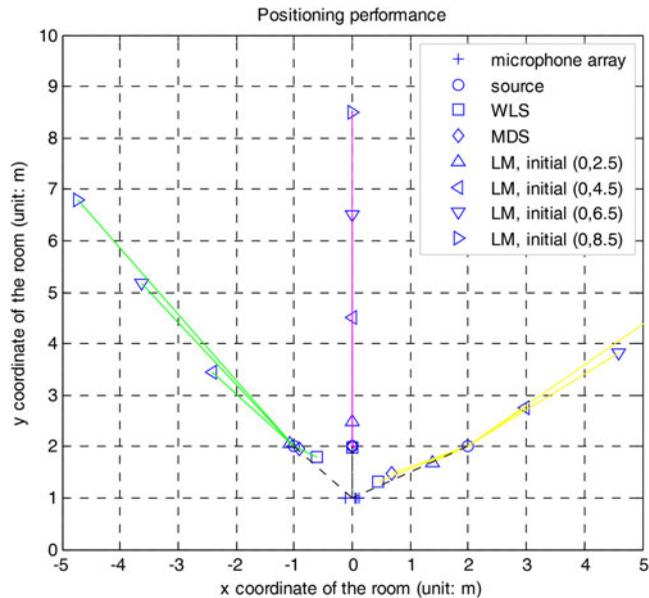


Figure 4. Localisation performance for sources at (0,2), (-1,2), and (2,2).

STD of 1 mm for the range estimation. The localisation performances of the three algorithms are shown in Figures 4 to 6.

From Figures 4–6, we conclude that:

- 1) The WLS method and the MDS method can give accurate estimates on source locations when the sound source is close to the Y-axis. The smaller the Y coordinate of the source is, the more accurate the estimated location is. The accuracy of the MDS method is better than that of the WLS method.
- 2) The LM method can provide accurate estimated directions of the sources. The estimated direction of the source is more accurate when the initial point lies closer to the centre of the microphone array. However, in the LM method, an accurate estimation of the exact coordinates of the source cannot be guaranteed, since the final result of the LM method depends on the choice of the initial point.

In conclusion, the WLS method and the MDS method perform well only for sources located close to the perpendicular bisector of the microphone array, while the LM method is particularly good at estimating the direction of the source. We should re-emphasise that, in contrast to the direction-finding algorithms like MUSIC which work only in the far-field condition, the LM method works in both far-field and near-field conditions.

Secondly, we evaluate the performance of the proposed hybrid localisation scheme (Algorithm 4). The details of the hybrid scheme are explained as follows. In Step 1 of Algorithm 4, the initial points for LM method are set to be (-2,2), (-1,3), (1,3), and (2,2). The estimated direction of the source is set to be the average of the four estimated directions calculated from the four initial points. In Step 3 of Algorithm 4, the MDS

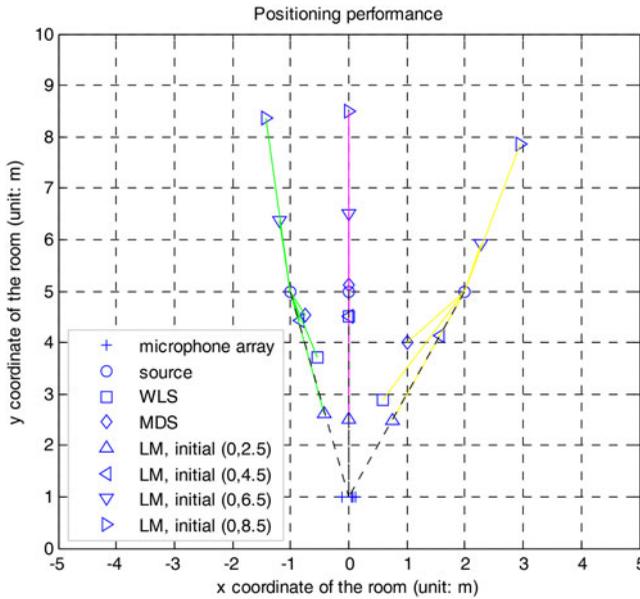


Figure 5. Localisation performance for sources at (0,5), (-1,5), and (2,5).

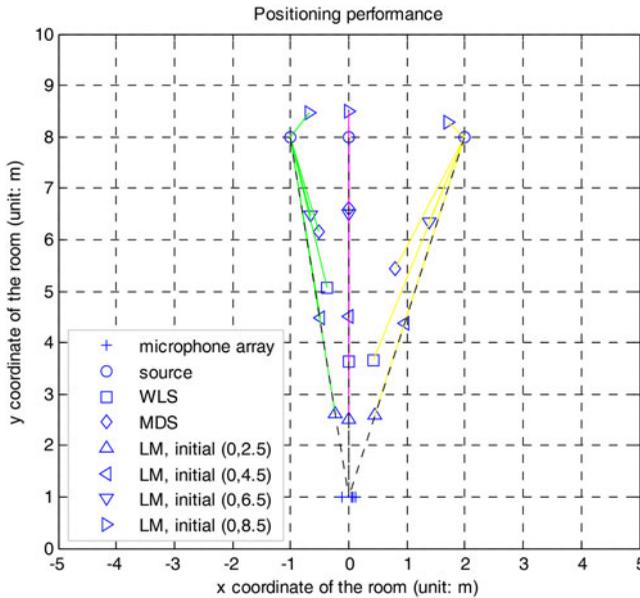


Figure 6. Localisation performance for sources at (0,8), (-1,8), and (2,8).

method is chosen to estimate the location of the source in the rotated coordinate system, since it performs a little better than the WLS method. The other simulation parameters are the same as those for Figures 4 to 6.

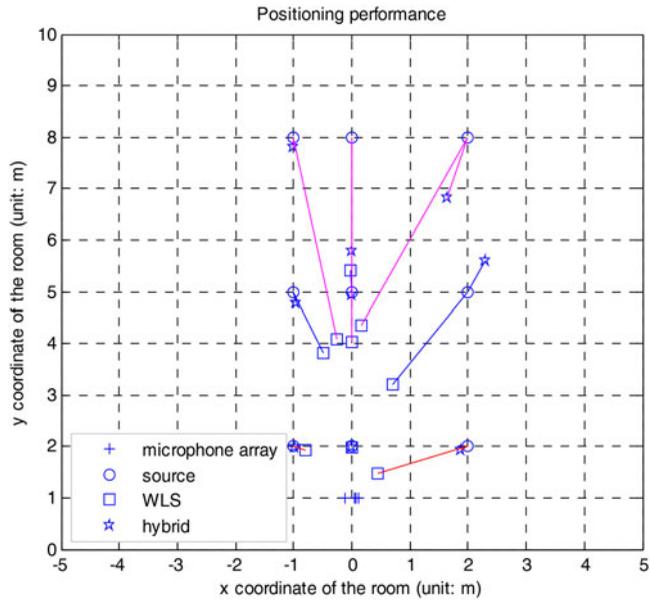


Figure 7. Localisation performance for sources at (0,2), (-1,2), (2,2), (0,5), (-1,5), (2,5), (0,8), (-1,8), (2,8).

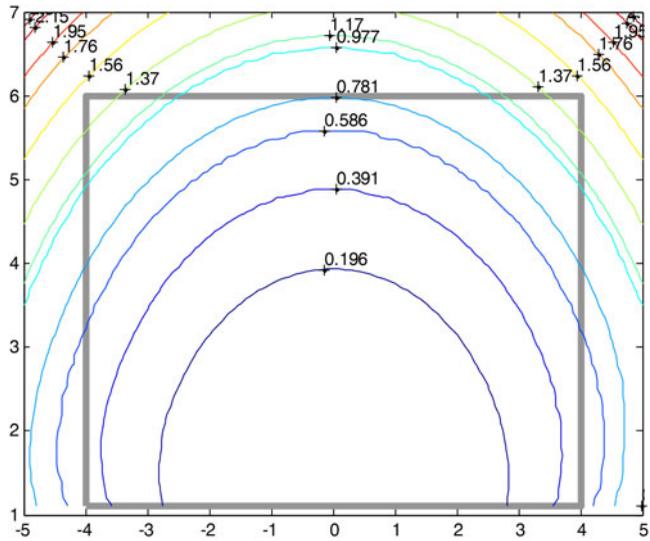


Figure 8. Estimate error distribution of the hybrid scheme in a target area.

In Figure 7, we compare the simulation results for the WLS method and the proposed hybrid scheme. Here, we choose the WLS method as benchmark because it is a well-known TDOA-based localisation algorithm and it has a similar performance as the MDS method as can be seen from Figures 4 to 6. Additionally, in Figure 8, we display the estimate error distribution of the hybrid scheme in a given area. The

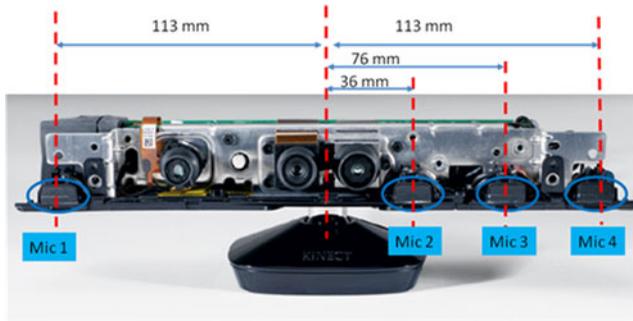


Figure 9. The Microphone array in a Kinect.

estimate error is defined as the distance between the true location and the estimated location of the same source.

From Figures 7 and 8, we summarise the following conclusions.

- 1) The estimated locations by the hybrid scheme are more accurate than those by the WLS method, especially for sources located relatively far away from the Y-axis.
- 2) The accuracy of the estimated locations by the hybrid scheme decreases when the source goes away from the centre of the microphone array.
- 3) The hybrid method could achieve an average accuracy of 0.32 m in the square area marked in grey in Figure 8.

5.2. Field tests. We verify the effectiveness of the investigated localisation algorithms using field tests. A Kinect device (Webb and Ashley, 2012) with four microphones arranged in a line has been used in our tests. The Kinect microphone array features four microphone capsules with each channel processing 16-bit audio at a sampling rate of 16 kHz. The dimensions of the Kinect microphone array are shown in Figure 9. The distance between the left most and the right most microphones is 0.226 m. Only one microphone is placed on the left-hand side and the three other microphones are located on the right-hand side of the device. The Kinect microphone array was originally designed for beam forming (Thomas et al., 2012). Previous research on the Kinect microphone array mainly focuses on the direction detection of a sound source instead of the exact location. In this paper, we use the Kinect for locating a sound source and estimating its coordinates.

Figure 10 shows the test setup, where a Kinect device is placed on a table near the wall and connected with a laptop via USB connection. In the test environment, as shown in Figure 11, a sound source is located on another table with the same height as the one that the Kinect device stands on. The local coordinate system defined for our tests is depicted in Figure 11.

Figure 12 shows a test example in which a sound source emits a signal four times, and the signal is recorded by the four channels of a Kinect microphone array. Each channel handles the signals of one microphone. The signals are utilised for time delay estimation using the algorithms described in Section 3. It should be noted that we conducted our field tests in a simple way in order to concentrate on verifying the effectiveness of the algorithms. The more sophisticated acoustic signal processing is beyond our scope and remains for future studies.

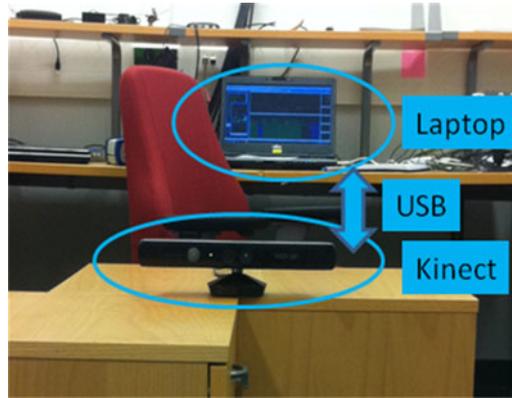


Figure 10. Test setup.

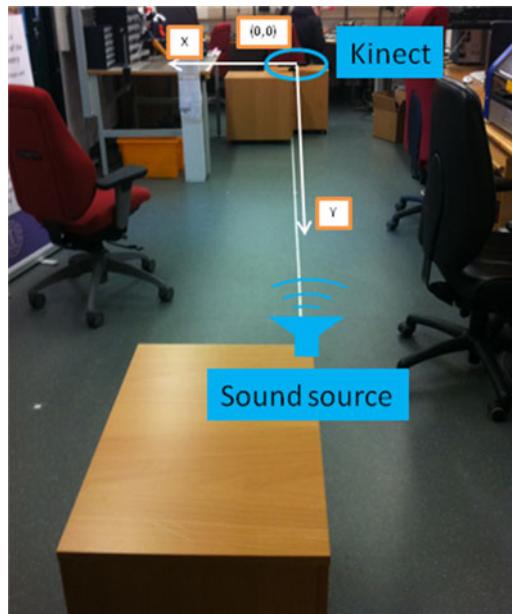


Figure 11. Test environment.

If we consider all the possible microphone pairs, i.e., (Mic1, Mic2), (Mic1, Mic3), (Mic1, Mic4), (Mic2, Mic3), (Mic2, Mic4), (Mic3, Mic4), an example of the delay estimates is shown in [Figure 13](#) and [Table 1](#), where the source is located at (1,1.9). From [Figure 13](#), we can see that both of the methods give good estimates of the true time delay, while the time domain delay estimate method is a little better than the GCC-PHAT delay estimate method. The maximum error of delay estimate is about 0.0066 ms, which is interpreted into about 2.26 mm error in the range measurement. The average deviation between the range measurement and the true range is 1.01 mm.

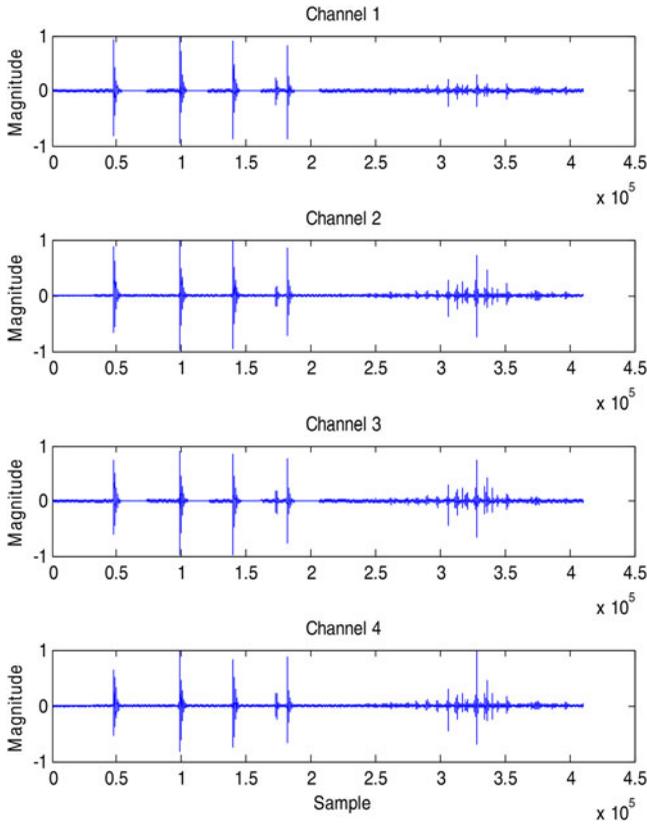


Figure 12. Signals of Kinect's four channels.

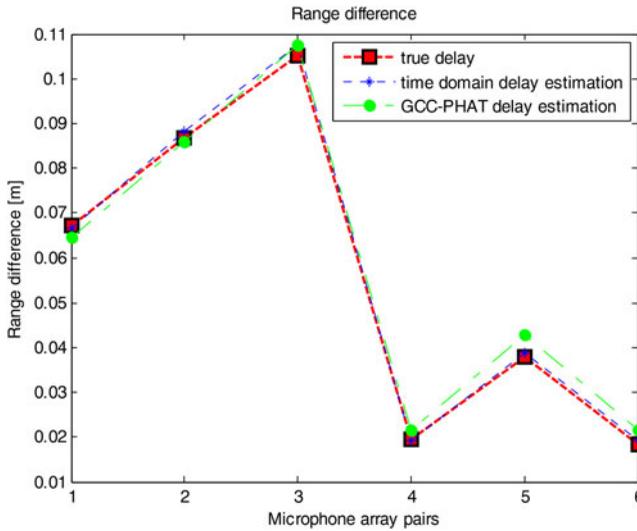


Figure 13. Time delay estimation experiment.

Table 1. Time Delay Estimation Experiment.

| Microphone pair index | Microphone pair | True delay (ms) | Time domain delay estimation (ms) | GCC-PHAT delay estimation (ms) |
|-----------------------|-----------------|-----------------|-----------------------------------|--------------------------------|
| 1 | (Mic1, Mic2) | 0.1956 | 0.1937 | 0.1875 |
| 2 | (Mic1, Mic3) | 0.2522 | 0.2562 | 0.2500 |
| 3 | (Mic1, Mic4) | 0.3059 | 0.3125 | 0.3125 |
| 4 | (Mic2, Mic3) | 0.0565 | 0.0562 | 0.0625 |
| 5 | (Mic2, Mic4) | 0.1103 | 0.1125 | 0.1250 |
| 6 | (Mic3, Mic4) | 0.0537 | 0.0562 | 0.0625 |

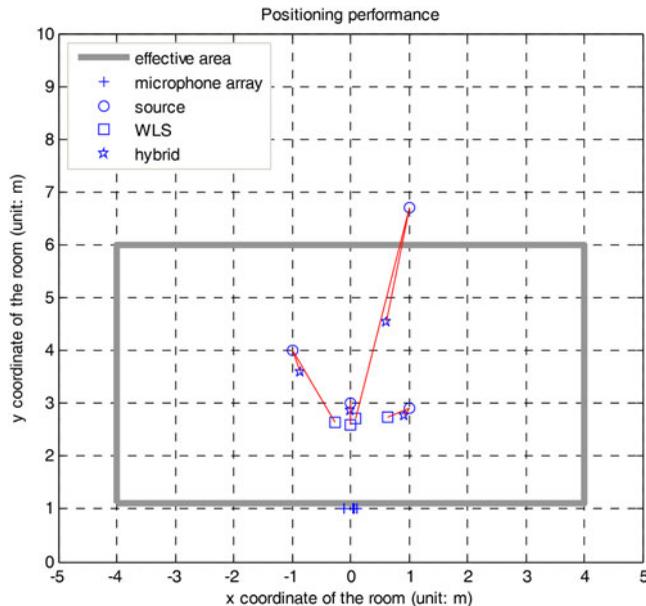


Figure 14. Localisation performance for sources located at (0,3), (1,2.9), (-1,4), and (1,6.7).

Table 2. Localisation Error for sources located at (0,3), (1,2.9), (-1,4), and (1,6.7).

| Source location | Localisation error of WLS (m) | Localisation error of hybrid (m) |
|-----------------|-------------------------------|----------------------------------|
| (0,3) | 0.4109 | 0.1355 |
| (1,2.9) | 0.3950 | 0.1554 |
| (-1,4) | 1.5444 | 0.4251 |
| (1,6.7) | 4.0794 | 2.1981 |

The coordinate system of the field test is the same as that used in Section 5.1. In Figure 14 and Table 2 we present the field test results of the hybrid scheme for another four sources located at (0,3), (1,2.9), (-1,4), and (1,6.7). The well-known WLS method is chosen as the benchmark.

From Figure 14 and Table 2 we can see that the estimated locations by the hybrid scheme are more accurate than those of the WLS method. In the same square grey

area plotted in Figure 8, the WLS method achieves an average error of 0.7834 m. The mean error of the proposed hybrid scheme is only 0.2387 m. Thus, our field tests verify the conclusions drawn in Section 5.1.

6. CONCLUSION AND FUTURE WORKS. Microphone arrays have been widely used for sound localisation. Small-scale microphone arrays have been deployed in consumer electronic devices such as Kinect and smartphones. This paper studies the possibility of using a small-scale linear microphone array for the purpose of locating a sound source. Both the time domain delay estimate method and the GCC-PHAT delay estimate method are applied to estimate the time difference of arrival in a microphone pair. Given the time difference estimations, we adapt the WLS method, the MDS method, and the LM method to calculate the coordinates of the sound source. Furthermore, we develop a hybrid scheme based on the above basic methods. Both simulations and field tests are conducted. Simulation results show that the proposed hybrid scheme outperforms the existing methods, which can achieve an average accuracy of 0.32 m in a room of 8 m by 5 m. Field tests are executed in a laboratory environment using a linear array in a Kinect device with a length of 0.226 m. The time delay estimate method is verified to be able to give delay estimates with an error less than 0.0066 milliseconds, which satisfies the requirement of localisation algorithms. Field test results corroborate the conclusions drawn from simulations.

The scope of this paper is aimed at localisation solutions. Therefore, the efficiency of the utilised algorithms is ignored. In the future, we will further study these topics, for instance, using the geometry information to constrain the iterative optimisation and restrict the search space. This paper only presents the test results for a static environment. In the future, the effects of obstacles, e.g., people, furniture, equipment etc., and tracking of moving target will be studied. Finally, we will extend our work to three-dimensional scenarios in future studies.

ACKNOWLEDGEMENTS

The authors are grateful to Dr. Hongmei Hu from the Medical Physics Section of Carl von Ossietzky Universität Oldenburg, Germany for her suggestions on acoustic signal processing. The authors also would like to thank Anttoni Jaakkola from the Finnish Geodetic Institute for providing the test device. The first author would like to thank Dr. Ming Ding from National ICT Australia (NICTA) for his helpful comments.

FINANCIAL SUPPORT

This work is jointly funded by China Satellite Navigation Office and the Science and Technology Commission of Shanghai Municipality. The funding project number is BDZX005. This work is also supported by fund from the Science and Technology Commission of Shanghai Municipality (No.14511100300), the Natural Science Foundation of China (No.6140050267), and partly sponsored by Shanghai Pujiang Program (No. 14PJ1405000).

REFERENCES

- Bahl, P. and Padmanabhan, V.N. (2000). Radar: An In-Building RF Based User Location and Tracking System. *Proceedings of Infocom—Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies*, Tel-Aviv, Israel.
- Brandstein, M.S. (1995). A framework for speech source localisation using sensor arrays.
- Brandstein, M. and Ward, D. (2001). *Microphone Arrays*. Springer.
- Carter, G.C. (1987). Coherence and time delay estimation. *Proceedings of the IEEE*, **75**(2), 236–255.
- Chan, Y.T. and Ho, K.C. (1994). A simple and efficient estimator for hyperbolic location. *IEEE Transactions on Signal Processing*, **42**(8), 1905–1915.
- Chen, W., Chen, R., Chen, Y., Kuusniemi, H., Wang, J. and Fu, Z. (2010). An Effective Pedestrian Dead Reckoning Algorithm Using a Unified Heading Error Model. *Proceedings of IEEE/ION PLANS 2010 Conference*, Indian Wells: Palm Springs, CA.
- Chen, R., Chen, W., Chen, X., Zhang, X. and Chen, Y. (2011a). Sensing Strides Using EMG Signal for Pedestrian Navigation. *GPS Solutions*, **15**(2), 161–170.
- Chen, R., Pei, L. and Chen, Y. (2011b). A Smart Phone Based PDR Solution for Indoor Navigation. *ION GNSS 2011*, Portland, OR.
- Chen, L., Yang, L. and Chen, R. (2012). Time delay tracking for positioning in DTV networks. *Proceedings of Ubiquitous Positioning, Indoor Navigation, and Location Based Service*, Helsinki, Finland.
- Chen, L., Pei, L., Kuusniemi, H., Chen, Y., Kröger, T. and Chen, R. (2013). Bayesian fusion for indoor positioning using bluetooth fingerprints. *Wireless personal communications*, **70**(4), 1735–1745.
- DiBiase, J.H., Silverman, H.F. and Brandstein, M.S. (2001). Robust localisation in reverberant rooms. *Microphone Arrays*, Berlin, Germany: Springer Berlin Heidelberg, 157–180.
- Filonenko, V., Cullen, C. and Carswell, J. (2010). Investigating ultrasonic positioning on mobile phones. *International Conference on Indoor Positioning and Indoor Navigation*, Zurich, Switzerland.
- Filonenko, V., Cullen, C. and Carswell, J.D. (2012). Asynchronous ultrasonic trilateration for indoor positioning of mobile phones. *Web and Wireless Geographical Information Systems*. Berlin, Germany.
- Foxlin, E. (2005). Pedestrian Tracking with Shoe-Mounted Inertial Sensors. *IEEE Computer Graphics and Applications*, **25**(6), 38–46.
- Gomes, G. and Sarmiento, H. (2009). Indoor Location System Using ZigBee Technology. *Proceedings of Third International Conference on Sensor Technologies and Applications*, Athens/Glyfada, Greece.
- Hoflinger, F., Zhang, R., Hoppe, J., Bannoura, A., Reindl, L.M., Wendeberg, J. and Schindelbauer, C. (2012). Acoustic Self-calibrating System for Indoor Smartphone Tracking. *International Conference on Indoor Positioning and Indoor Navigation*, Sydney, NSW.
- Holm, S. (2012). Ultrasound positioning based on time-of-flight and signal strength. *International Conference on Indoor Positioning and Indoor Navigation*, Sydney, NSW.
- Ianniello, J. (1982). Time delay estimation via cross-correlation in the presence of large estimation errors. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **30**(6), 998–1003.
- IndoorAtlas Ltd. (2011). *Game-changing indoor location service*. <http://www.indooratlas.com>. Accessed October 2014.
- Janson, T., Schindelbauer, C. and Wendeberg, J. (2010). Self-localisation application for iPhone using only ambient sound signals. *International Conference on Indoor Positioning and Indoor Navigation*. Zurich, Switzerland.
- Khalil, F., Jullien, J.P. and Gilloire, A. (1994). Microphone array for sound pickup in teleconference systems. *Journal of the Audio Engineering Society*, **42**(9), 691–700.
- Knapp, C. H. and Carter, G. C. (1976). The generalized correlation method for estimation of time delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, **24**(4), 320–332.
- Kuusniemi, H., Liu, J., Pei, L., Chen, Y., Chen, L. and Chen, R. (2012). Reliability considerations of multi-sensor multi-network pedestrian navigation. *IET Radar, Sonar & Navigation*, **6**(3), 157–164.
- Liu, J., Chen, R., Pei, L., Chen, W., Tenhunen, T., Kuusniemi, H., Kröger, T. and Chen, Y. (2010). Accelerometer Assisted Wireless Signals Robust Positioning Based on Hidden Markov Model. *Proceedings of the IEEE/ION Position, Location and Navigation Symposium 2010*, Indian Wells, CA.
- Liu, J., Chen, R., Chen, Y., Pei, L. and Chen, L. (2012a). iParking: An Intelligent Indoor Location-Based Smartphone Parking Service. *Sensors*, **12**(11), 14612–14629.
- Liu, J., Chen, R., Pei, L., Guinness, R. and Kuusniemi, H. (2012b). Hybrid smartphone indoor positioning solution for mobile LBS. *Sensors*, **12**(12), 17208–17233.
- Lourakis, M.I. (2005). A brief description of the Levenberg-Marquardt algorithm implemented by levmar. *Institute of Computer Science, Foundation for Research and Technology*, **4**, 1–6.

- Mathews, M.B., Maccoran, P.F., Gold, K.L. (2011). SCP Enabled Navigation Using Signals of Opportunity in GPS Obstructed Environments. *Navigation*, **58**(2), 91–110.
- Michel, U. and Barsikow, B. (2003). Localisation of sound sources on moving vehicles with microphone arrays. *Proceedings 5th European Conference on Noise Control Euronoise 2003*, Naples, Italy.
- Moré, J.J. (1978). The Levenberg-Marquardt algorithm: implementation and theory. *Numerical analysis*, Berlin, Germany: Springer Berlin Heidelberg, 105–116.
- Ni, L.M., Liu, Y., Lau, Y.C. and Patil, A.P. (2004). Landmarc: Indoor location sensing using active RFID. *Wireless networks*, **10**(6), 701–710.
- Nocedal, J. and Wright, S. (1999). *Numerical Optimization*. Springer-Verlag.
- Pahlavan, K., Akgul, O.F., Heidari, M., Hatami, A., Elwell, M.J. and Tingley, D.R. (2006). Indoor Geolocation in the Absence of Direct Path. *IEEE Wireless Communication*, **13**(6), 50–58.
- Pei, L., Chen, R., Liu, J., Chen, W., Kuusniemi, H., Tenhunen, T., Kröger, T., Chen, Y., Leppäkoski, H. and Takala, J. (2010a). Motion Recognition Assisted Indoor Wireless Navigation on a Mobile Phone. *Proceedings of the 23rd International Technical Meeting of the Satellite Division of the Institute of Navigation*, Portland, OR.
- Pei, L., Chen, R., Liu, J., Kuusniemi, H., Tenhunen, T. and Chen, Y. (2010b). Using inquiry-based Bluetooth RSSI probability distributions for indoor positioning. *Journal of Global Positioning Systems*, **9**(2), 122–130.
- Pei, L., Chen, R., Liu, J., Tenhunen, T., Kuusniemi, H. and Chen, Y. (2010c). Inquiry-Based Bluetooth Indoor Positioning via RSSI Probability Distributions. *Proceedings of the Second International Conference on Advances in Satellite and Space Communications*, Athens, Greece.
- Pei, L., Chen, R., Liu, J., Kuusniemi, H., Chen, Y. and Tenhunen, T. (2011). Using Motion-Awareness for the 3d Indoor Personal Navigation on a Smartphone. *Proceedings of the 24rd International Technical Meeting of the Satellite Division of the Institute of Navigation*, Portland, OR.
- Pei, L., Liu, J., Guinness, R., Chen, Y., Kuusniemi, H. and Chen, R. (2012). Using LS-SVM based motion recognition for smartphone indoor wireless positioning. *Sensors*, **12**(5), 6155–6175.
- Pei, L., Guinness, R., Chen, R., Liu, J., Kuusniemi, H., Chen, Y., Chen, L. and Kaistinen, J. (2013). Human Behavior Cognition Using Smartphone Sensors. *Sensors*, **13**(2), 1402–1424.
- Pertila, P., Mieskolainen, M. and Hamalainen, M.S. (2012). Passive self-localisation of microphones using ambient sounds. *Signal Processing Conference, 2012 Proceedings of the 20th European*. Bucharest, Romania.
- Piet, J., Michel, U. and Böhning, P. (2002). Localisation of the acoustic sources of the A340 with a large microphone array during flight tests. *8th AIAA/CEAS Aeroacoustics Conference*, Breckenridge, Colorado.
- Priyantha, N.B., Chakraborty, A. and Balakrishnan, H. (2000). The Cricket Location-Support System. *Proceedings of the 6th Annual International Conference on Mobile Computing and Networking*, Boston, MA.
- Ruotsalainen, L., Kuusniemi, H. and Chen, R. (2011). Visual-aided Two-dimensional pedestrian indoor navigation with a smartphone. *Journal of Global Positioning Systems*, **10**(1), 11–18.
- Schulte-Werning, B., Jäger, K., Strube, R. and Willenbrink, L. (2003). Recent developments in noise research at Deutsche Bahn (noise assessment, noise source localisation and specially monitored track). *Journal of Sound and Vibration*, **267**(3), 689–699.
- Steinberg, B. D. (1976). *Principles of aperture and array system design: Including random and adaptive arrays*. Wiley-Interscience.
- Storms, W., Shockley, J. and Raquet, J. (2010). Magnetic field navigation in an indoor environment. *Proceedings of Ubiquitous Positioning Indoor Navigation and Location Based Service*, Kirkkonummi, Finland.
- Syrjärinne, J. (2001). Studies on Modern Techniques for Personal Positioning. Ph.D. thesis, Tampere University of Technology, Tampere, Finland.
- Thomas, M.R., Ahrens, J. and Tashev, I. (2012). Optimal 3D beam forming using measured microphone directivity patterns. *Proceedings of International Workshop on Acoustic Signal Enhancement*, Aachen, Germany.
- Valenzise, G., Gerosa, L., Tagliasacchi, M., Antonacci, F. and Sarti, A. (2007). Scream and gunshot detection and localisation for audio-surveillance systems. *IEEE Conference on Advanced Video and Signal Based Surveillance*, London, England.
- Webb, J. and Ashley, J. (2012). *Beginning Kinect Programming with the Microsoft Kinect SDK*. Apress.
- Wei, H.W., Wan, Q., Chen, Z.X. and Ye, S.F. (2008). Multidimensional scaling-based passive emitter localisation from range-difference measurements. *Signal Processing, IET*, **2**(4), 415–423.
- Wei, H.W., Peng, R., Wan, Q. and Chen, Z.X. (2010). Multidimensional scaling analysis for passive moving target localisation with TDOA and FDOA measurements. *IEEE Transactions on Signal Processing*, **58**(3), 1677–1688.

APPENDIX

The detail of Step 6 of Algorithm 3 is as follows:

Solving the minimization problem in Step 6 of Algorithm 3

6.1: $\tilde{\mathbf{J}} = \mathbf{J}(\mathbf{x}_s^{(l)}) \cdot \mathbf{D}^{-1}$.

6.2: $\mathbf{g} = \tilde{\mathbf{J}}^T \cdot \boldsymbol{\varepsilon}(\mathbf{x}_s^{(l)})$.

6.3: $\mathbf{B} = \tilde{\mathbf{J}}^T \cdot \tilde{\mathbf{J}}$.

6.4: Eigenvalue decomposition: $\mathbf{B} = \mathbf{Q}\boldsymbol{\Lambda}\mathbf{Q}^T$,

where $\boldsymbol{\Lambda} = \text{diag}\{\lambda_1, \lambda_2\}$, $\lambda_1 \leq \lambda_2$, $\mathbf{Q}\mathbf{Q}^T = \mathbf{Q}^T\mathbf{Q} = \mathbf{I}$

6.5: If $\lambda_1 \leq 0$, then go to Step 6.7

6.6: If $\lambda_1 > 0$, then let $\mathbf{p} = -\mathbf{Q}\boldsymbol{\Lambda}^{-1}\mathbf{Q}^T\mathbf{g}$. If $\|\mathbf{p}\| < \Delta_l$, then go to Step 6.20; if $\|\mathbf{p}\| \geq \Delta_l$, then go to Step 6.7.

6.7: Initialise $\lambda^{(n)} = \lambda^{(0)}$ which satisfies $\lambda^{(0)} > -\lambda_1$.

6.8: QR decomposition: $\begin{pmatrix} \tilde{\mathbf{J}} \\ \sqrt{\lambda^{(n)}}\mathbf{I} \end{pmatrix} = \mathbf{Q}_\lambda \begin{pmatrix} \mathbf{R}_\lambda \\ \mathbf{0} \end{pmatrix}$,

where \mathbf{Q}_λ is orthogonal, \mathbf{R}_λ is upper triangular.

6.9: Solve $\mathbf{p}(\lambda^{(n)})$ from $\mathbf{R}_\lambda^T \mathbf{R}_\lambda \mathbf{p} = -\mathbf{g}$.

6.10: If $\|\|\mathbf{p}\| - \Delta_l\| < \text{th}_\lambda$, where th_λ is a preset threshold, then stop iteration and go to Step 6.20.

6.11: If $\|\|\mathbf{p}\| - \Delta_l\| \geq \text{th}_\lambda$

6.12: Solve \mathbf{q} from $\mathbf{R}_\lambda^T \mathbf{q} = \mathbf{p}$.

6.13: Set $h = 1$

6.14: $\lambda^{(n+1)} = \lambda^{(n)} + h \cdot \frac{\|\mathbf{p}\|^2}{\|\mathbf{q}\|^2} \cdot \frac{\|\mathbf{p}\| - \Delta_l}{\Delta_l}$.

6.15: QR decomposition: $\begin{pmatrix} \tilde{\mathbf{J}} \\ \sqrt{\lambda^{(n+1)}}\mathbf{I} \end{pmatrix} = \mathbf{Q}_\lambda \begin{pmatrix} \mathbf{R}_\lambda \\ \mathbf{0} \end{pmatrix}$,

where \mathbf{Q}_λ is orthogonal, \mathbf{R}_λ is upper triangular.

6.16: Solve $\mathbf{p}(\lambda^{(n+1)})$ from $\mathbf{R}_\lambda^T \mathbf{R}_\lambda \mathbf{p} = -\mathbf{g}$.

6.17: If $\|\mathbf{p}(\lambda^{(n+1)})\| < \|\mathbf{p}(\lambda^{(n)})\|$ and $\lambda^{(n+1)} > 0$, then $n \leftarrow n + 1$, go to Step 6.8.

6.18: If $\|\mathbf{p}(\lambda^{(n+1)})\| < \|\mathbf{p}(\lambda^{(n)})\|$ and $\lambda^{(n+1)} \leq 0$, then let $\lambda^{(n+1)} = \lambda^{(n)}/2$, go to Step 6.8.

6.19: If $\|\mathbf{p}(\lambda^{(n+1)})\| \geq \|\mathbf{p}(\lambda^{(n)})\|$, then let $h \leftarrow h/2$, go to Step 6.14.

6.20: $\mathbf{p}_l = \mathbf{D}^{-1}\mathbf{p}$.