PANDAA: Physical Arrangement Detection of Networked Devices through Ambient-Sound Awareness

[1]Zheng Sun, [1]Aveek Purohit, [2]Kaifei Chen, [2]Shijia Pan, [3]Trevor Pering, [1]Pei Zhang
[1]Electrical and Computer Engineering, Carnegie Mellon University
[2]Computer Science and Technology, University of Science and Technology of China
[3]Independent
{zhengs, apurohit, peizhang}@cmu.edu, {kaifei.chen, shijia.pan, grafnu}@gmail.com

ABSTRACT
Future ubiquitous home environments can contain 10s or 100s of devices. Ubiquitous services running on these devices (i.e. localizing users, routing, security algorithms) will commonly require an accurate location of each device. In order to obtain these locations, existing techniques require either a manual survey, active sound sources, or estimation using wireless radios. These techniques, however, need additional hardware capabilities and are intrusive to the user. Non-intrusive, automatic localization of ubiquitous computing devices in the home has the potential to greatly facilitate device deployments.

This paper presents the PANDAA system, a zero-configuration spatial localization system for networked devices based on ambient sound sensing. After initial placement of the devices, ambient sounds, such as human speech, music, footsteps, finger snaps, hand claps, or coughs and sneezes, are used to autonomously resolve the spatial relative arrangement of devices using trigonometric bounds and successive approximation. Using only time difference of arrival measurements as a bound for successive estimations, PANDAA is able to achieve an average of 0.17 meter accuracy for device location in the meeting room deployment.

ACM Classification Keywords
C.3 Special-purpose and application-based systems: Signal processing systems.

General Terms
Algorithms, Design, Experimentation.

Author Keywords
Arrangement detection, networked devices, localization.

INTRODUCTION
PANDAA (Physical Arrangement Detection of Networked Devices through Ambient-Sound Awareness) is a spatial localization technique that uses ambient sounds. By forming a collective understanding across a set of nodes of the time-of-arrival for ambient audio events, PANDAA can automatically determine the physical arrangement of nodes without prior calibration of their locations. After initial deployment, ambient sounds, such as human speech, music, footsteps, finger snaps, hand claps, or coughs and sneezes, are used to autonomously resolve the spatial relative arrangement of the devices using trigonometric bounds and successive approximation.

Automatic localization of ubiquitous computing devices has the potential to greatly facilitate the introduction of technology into consumer or other "non administrated" environments. For example, consider the simple case of a consumer purchasing a new smart television for their home. Manual configuration would entail the user specifying which room the device is in, information that should be easily discernible by the system itself. Consider, then, the difficulty in configuring a ubiquitous computing environment that contains 10s or 100s of devices. As the number of devices increases, so does the need for very-low overhead configuration mechanisms. Fortunately, an increase in the number of devices also increases the potential for collaborative techniques to handle configuration.

Once devices are accurately positioned, the system can also be used to localize the source of audio events themselves. This can enable the automatic deployment of a variety of context-aware applications, such as audio surveillance [11], speaker localization [10], activity detection [5], and patient and elderly monitoring [25]. In these applications, locations of audio events being monitored are of great importance causing a pronounced "chicken-and-egg" problem where sensing can be used to determine the location of ambient acoustic events, but only after the location of the sensors is known.

The PANDAA system leverages microphones that already exist in various consumer devices, such as laptops, tablets, mobile phones, home theaters, smart TVs, etc. One single microphone per device is used to detect usable segments of ambient sound generated in a room. Next, the time difference of sound arrival (TDoA) between devices is calculated and used to iteratively estimate inter-device distances. These distances are then used to determine the overall arrangement of devices. Finally multiple TDoA measurements are combined to improve arrangement detection accuracy over time.

The implementation and evaluation of the PANDAA system

Permission to make digital or hard copies of all or part of this work for personal or classroom use is granted without fee provided that copies are not made or distributed for profit or commercial advantage and that copies bear this notice and the full citation on the first page. To copy otherwise, or republish, to post on servers or to redistribute to lists, requires prior specific permission and/or a fee.
UbiComp'11, September 17–21, 2011, Beijing, China.
Copyright 2011 ACM 978-1-4503-0630-0/11/09...$10.00.

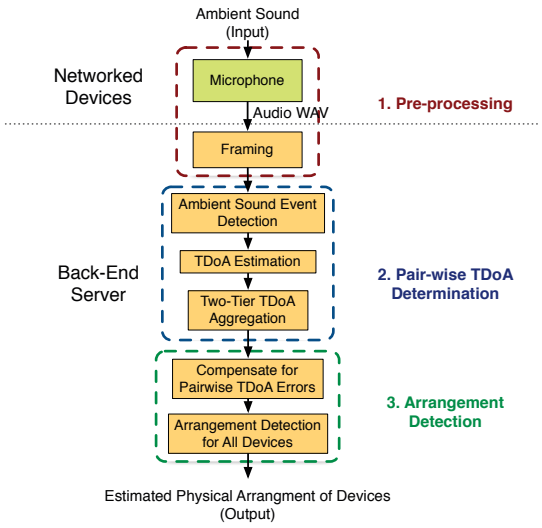

Figure 1. The system architecture of PANDAA.

shows that calibration-free device localization is possible. To cover broad ranges of ambient sounds, we tested the system under different distribution of ambient sound locations with cough, speech, and two types of music that commonly exist in office and home environments. To deal with time-varying indoor noises and compensate for reflections from walls and obstacles, we tested how sound event detection, aggregation and noise modeling techniques can be applied to improve system performance. We show that PANDAA can achieve 0.17m arrangement detection accuracy in real indoor environments. And with such accuracy, the system is able to localize sound sources to within 0.5m 80% of the time, yielding an error level that is comparable to previous work that used manual surveys [10, 25].

The key contributions of this paper are as follows:

1. We provide a proof of concept for a zero-configuration spatial localization of nodes based on ambient sound sensing.
2. We design a set of algorithms that leverages multi-node collaboration to compensate for indoor ambient sound effects, such as echos and ambient noise.
3. We design an algorithm that iteratively determines and improves the relative arrangement of the sensing nodes over time.

The following sections provide a detailed description of the PANDAA system, along with experimental evaluations to show its efficacy. Then we highlight the difference between related work and PANDAA. Finally, we conclude our work and summarize our contributions.

THE PANDAA SYSTEM

The system architecture of PANDAA is shown in Figure 1. Each sensor node is assumed to have a microphone that captures ambient sound, a microprocessor for basic signal processing, and a radio for communicating results back to a central server. In this paper, we assume all devices in the room are placed in a 2-D plane. However, the algorithms presented can be easily extended to cover 3-D cases. No further assumptions are made about the 2-D physical location of the devices and ambient sound sources. The devices themselves are assumed to be wirelessly connected to the central server. Additionally, it is assumed that the devices themselves do not move. Time synchronization is achieved over the local wireless network.

The PANDAA system addresses the major challenges of ambient sound-based arrangement detection as follows.

1. **Choosing Usable Ambient Sound Segments.** Ambient sounds, such as music played on a radio, human speech, noise from a working vacuum cleaner or a barking dog, may vary significantly in signal-to-noise ratio (SNR). In addition, varying proximity to the sound source can lead to significant difference in SNR. PANDAA addresses this challenge using an algorithm that can automatically detect *impulsive sounds*. Impulsive sounds are short duration sounds with relatively higher amplitude, such as human cough, finger snaps, or beats in a song.
2. **Correcting Inaccurate TDoA Measurements.** In indoor environments, TDoA measurements can be affected by environmental factors, such as reflections, non-line-of-sight (LoS) path, or ambient noise. These effects are location-dependent and time-variant. Consider a resident walking in the room. His/her changing location may temporarily cause a few devices to lose LoS, which can cause erroneous TDoA measurements. To compensate for TDoA errors, PANDAA uses a novel two-tier TDoA aggregation algorithm that identifies sounds originating from the same sound source and averages TDoA measurements over them.
3. **Localizing Devices From TDoA Measurements.** TDoA measurements from one single sound event are insufficient for estimating distance between two devices. PANDAA addresses this challenge by considering TDoA measurements from multiple ambient sound sources over time, to estimate inter-device distances and iteratively improve accuracy.

The following sections describe in detail the algorithm design of the PANDAA system.

Choosing Usable Ambient Sound Segments

PANDAA uses a frame-based impulse detection approach to automatically detect impulsive sounds that are distinct and have high signal-to-noise ratios (SNR). This approach consists of two steps.

First, compensating for hardware variation among devices by eliminating each microphone's circuit noise. This is achieved by recording a period of ambient sound under quiet conditions, which is described in step two below.

Second, since circuit noise can be assumed independent to environmental sound levels, we calculate the root mean square (RMS) of framed audio signals as an indicator of current local loudness from the microphone. The minimum RMS κ_i^c of the recorded audio under quiet conditions is used to quantify circuit noise on device D_i. For each device pair D_i and D_j, the circuit noise levels $\kappa_{i(j)}^c$ are subtracted from the RMS values $\kappa_{i(j)}$ of each incoming frame on the two devices to obtain ambient sound levels. The results are then com-

Algorithm 1 Ambient Sound Event Detection

for each device pair D_i and D_j **do**
  Compute $\kappa_{i(j)}$ for the current frame on D_i and D_j
  $\kappa_{i(j)} \leftarrow \kappa_{i(j)} - \kappa^c_{i(j)}$
  if $InEvent =$ NO **then**
    if for either D_i or D_j, or both, $\kappa_{i(j)} \geq \kappa^{aver}_{i(j)} \cdot \alpha_\kappa$ **then**
      $InEvent \leftarrow$ YES // *A sound event starts*
      Save the current frame on D_i and D_j into buffer
    else
      $\kappa^{aver}_{i(j)} \leftarrow \frac{\kappa^{aver}_{i(j)} + \kappa_{i(j)}}{2}$
    end if
  else
    if for either D_i or D_j, $\kappa_{i(j)} \geq \kappa^{aver}_{i(j)} \cdot \alpha_\kappa$ **then**
      Save the current frame on D_i and D_j into buffer
    else
      $InEvent \leftarrow$ NO // *A sound event ends*
    end if
  end if
end for

pared with iteratively updated admission thresholds $\kappa^{aver}_{i(j)}$. If

$$\kappa_{i(j)} - \kappa^c_{i(j)} \geq \kappa^{aver}_{i(j)} \cdot \alpha_\kappa \quad (1)$$

holds for either D_i or D_j, or both, an impulsive sound is detected, indicating a sound event starts; otherwise, the frame is discard as ambient noise. A sound event ends when Eq. (1) does not hold for incoming frames on either device D_i or D_j.

The value $\kappa^{aver}_{i(j)}$ tracks the average level of ambient noise at each device, and is updated as a moving average of the current $\kappa^{aver}_{i(j)}$ and the RMS value of the latest discarded frame. α_κ is an RMS ratio between impulsive sounds and ambient noises.

Using this impulsive sound detection approach, PANDAA can automatically extract high SNR sound events from a variety of ambient sound types. We observe that a smaller α_κ reduces probability of missing usable sound events, but increases computational cost. During the experiments, we empirically set α_κ to 1.80 to reduce probability of missing distant impulsive sounds. However, an adaptive threshold could also be applied to further reduce computation. Algorithm 1 outlines the ambient sound event detection process.

Time Difference of Arrival Estimation

After detecting a sound event, each pair of devices use audio frames in the detected event to compute time difference of sound arrivals (TDoA).

We denote the distance between a sound source and the two devices as d_{si} and d_{sj}, then the TDoA of device D_i and D_j is computed as

$$\tau_{ij} = \frac{d_{si} - d_{sj}}{v} \quad (2)$$

where v is the speed of sound. TDoA can be estimated by maximizing the cross-correlation of received audio signals on the two devices. Due to its effectiveness for non-stationary signals in the presence of noise and echos, we choose the generalized cross-correlation (GCC) algorithm, which estimates TDoA by solving

$$\tau_{ij} = \arg\max_{\tau \in \mathcal{R}} \frac{1}{K} \sum_{k=0}^{K-1} \Psi_{ij}[k] C_{ij}[k] e^{jk\frac{2\pi}{k}\tau}, \quad (3)$$

where $\Psi_{ij}[k]$ is a weighting function, and $C_{ij}[k] = X_i[k]X'_j[k]$ is the cross-spectrum of audio frames on device D_i and D_j, respectively [7, 8]. K is the number of audio samples in one frame.

The solution to Eq. (3) can be found by using linear-search. One variant of GCC algorithm defines the weighting function using a *phase transform* (GCC-PHAT) [23], and performs considerably better than its counterpart version under an echo-rich indoor environment. The GCC-PHAT algorithm defines the weighting function as $\Psi^{phat}_{ij}[k] = \frac{1}{|C_{ij}[k]|}$.

To reduce effects from ambient noise, such as low and steady hum from air conditioning in the room, we apply a bandpass filter to suppress the energy components with frequencies lower than 300Hz or higher than 6KHz. The filter is defined as

$$\Psi^{bandpass}_{ij}[k] = \begin{cases} 1 & 300 \leq \frac{k}{K} \cdot f_s \leq 6000 \\ 0 & \text{otherwise} \end{cases}, \quad (4)$$

where f_s is the sampling rate of each microphone on the devices. Finally the entire weighting function is defined as $\Psi_{ij}[k] = \Psi^{phat}_{ij}[k] \times \Psi^{bandpass}_{ij}[k]$.

Correcting Inaccurate TDoA Measurements

To compensate for TDoA errors caused by reflections and ambient noises, PANDAA uses a two-tier TDoA aggregation algorithm.

Lower Tier: Since sound events are typically tens of milliseconds in length, the locations of sound sources can be assumed to be stationary within a single sound event [8]. Based on this assumption, the lower tier aggregates cross-spectrums $C_{ij}[k]$ over successive frames in the same sound event to suppress any frame-to-frame effects that are uncorrelated. First, each device pair D_i and D_j uses all frames in the current sound event to compute *aggregated* cross-spectrum

$$C_{ij}[k] = \frac{1}{N} \sum_{n=0}^{N-1} X_{i,n}[k] X'_{j,n}[k], \quad (5)$$

where N is the number of frames in the current event. Then, each pair uses Eq.(3) to compute TDoA for the current sound event.

Upper Tier: While the lower-tier aggregation is sufficient for reducing uncorrelated frame-level noise, longer lasting ambient effects, such as a moving person blocking acoustic LoS of several nodes, can also significantly alter the TDoA measurements. To handle such effects, that are uncorrelated between consecutive events, we design an upper-tier aggregation that averages TDoA estimates over multiple consecutive sound events belonging to the same sound source.

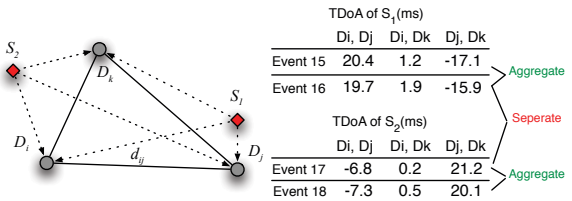

Figure 2. As locations of sound sources can be physically apart, TDoA measurements of consecutive sound events from the same source tend to be similar (such as the ones of Event 15 and 16), whereas those from different sources are dissimilar (such as the ones of Event 16 and 17). This information can be used to aggregate TDoA measurements from the same sound source and separate those from different sources.

The major challenge of the upper-tier aggregation is in separating sound events originating from different sound sources. Solitary devices cannot differentiate between multiple sound sources. However, as networked systems usually have multiple devices in a room, a group of networked devices can generate multiple TDoA measurements. For example, in Figure 2, TDoA measurements from the same sound source tend to be similar whereas those from different sources are more dissimilar. Considering TDoA measurements of multiple device pairs as a high-dimensional space, sound events from different sources can thus be separated by computing distances between TDoA measurements.

We note that there are two types of TDoA distances that should be considered. First, for the same sound source, different device pairs may experience different SNR measurements. According to the ambient sound detection algorithm, some of the device pairs may be able to detect the impulsive sounds whereas others may not. We consider this issue as the *hearability* difference between device pairs.

Second, if one device pair is able to make multiple TDoA measurements from the current sound source, it tends to get similar measurements with relatively small variation. We consider this as the TDoA *similarity*. Sound events from the same sound sources should be consistent in hearability and generate TDoA values with high similarity.

Based on this intuition, we define a hearability-vector to quantize the hearability difference between consecutive sound events, i.e.

$$\overrightarrow{\beta} = \{\beta_{1,2}, \beta_{1,3}, ...\beta_{2,3}, \beta_{2,4}, ...\}_{i,j}, \quad (6)$$

where $\beta_{i,j}$ equals 1 if a devise pair D_i, D_j is able to compute TDoA from the current sound event; otherwise 0. We also define a TDoA-vector to contain all the pairwise TDoA values for the current sound event, i.e.

$$\overrightarrow{\tau} = \{\tau_{1,2}, \tau_{1,3}, ...\tau_{2,3}, \tau_{2,4}, ...\}_{i,j}, \quad (7)$$

where $\tau_{i,j}$ is the TDoA value for the device pair D_i, D_j computed through the lower-tier aggregation. Both the hearability-vector and the TDoA-vector have $\frac{N_D(N_D-1)}{2}$ dimensions, where N_D is the total number of devices deployed in a room.

Then, the Hamming distance between the hearability-vectors of the current sound event, $\overrightarrow{\beta_{cur}}$, and that of the previous sound event, $\overrightarrow{\beta_{prev}}$, is computed to quantize consistency in hearability, i.e.

$$d_{\mathrm{H}} = \frac{\#(\overrightarrow{\beta_{cur}} \neq \overrightarrow{\beta_{prev}})}{\#\text{ of device pairs}}. \quad (8)$$

Meanwhile, the Euclidean distance between the TDoA-vectors of the current and previous sound events is computed to quantize the similarity in TDoA values.

Since difference in hearability may result in missing elements in the two TDoA-vectors, the common elements that *both* TDoA-vectors possess and which are valid, are first determined. Then, the Euclidean distance is computed *only* based on these common elements and finally normalized, i.e.

$$d_{\mathrm{E}} = \frac{\|\overrightarrow{\tau_{cur}} - \overrightarrow{\tau_{prev}}\|}{\#\text{ of common elements}}. \quad (9)$$

Finally, the per-dimensional hybrid distance is computed as the product of the two distances

$$d_{\mathrm{ALL}} = d_{\mathrm{H}} \cdot d_{\mathrm{E}}, \quad (10)$$

which is compared with a threshold d_{thr} to determine if the two consecutive sound events are considered as from the same source. If so, the TDoA values of the two events are averaged to suppress effect from ambient noise. Otherwise, the sound events are separated and treated as if they are from different sources. Algorithm 2 shows the two-tier TDoA aggregation process.

Algorithm 2 Two-Tier TDoA Aggregation

Lower-tier aggregation:
for each device pair D_i and D_j **do**
  Aggregate cross-spectrums within the current sound event according to Eq. (5).
  Compute TDoA values for the current sound event using Eq.(3).
end for

Upper-tier aggregation:
for each sound event **do**
  Create the hearability-vector $\overrightarrow{\beta_{cur}}$ for the current sound event according to Eq.(6); compute the Hamming distance d_{H} between $\overrightarrow{\beta_{cur}}$ and $\overrightarrow{\beta_{prev}}$ according to Eq.(8).
  Create the TDoA-vector $\overrightarrow{\tau_{cur}}$ using TDoA values of the current event according to Eq.(7).
  In the TDoA-vector, determine common elements in which both the current and previous events have valid TDoA values, and compute the per-dimensional Euclidean distance d_{E} between $\overrightarrow{\tau_{cur}}$ and $\overrightarrow{\tau_{prev}}$ according to Eq.(9).
  Compute hybrid distance $d_{\mathrm{ALL}} = d_{\mathrm{H}} \cdot d_{\mathrm{E}}$
  if $d_{\mathrm{ALL}} \leq d_{thr}$ **then**
    $\tau_{ij} = \frac{\tau_{ij}+\tau'_{ij}}{2}$, where τ'_{ij} is the TDoA value of device pair D_i,D_j for the previous sound events
  end if
end for

Localizing Devices From TDoA Measurements

To determine the physical arrangement of networked devices, PANDAA utilizes a two step approach. 1) obtain the pair-

wise TDoA values as an estimate of the lower bound of the inter-device distances, and 2) use the distribution of TDoA collected over time to estimate the true inter-device distances, from which device arrangement is derived.

Determine Pairwise Distance Using TDoA Noise Model

Consider two devices D_i and D_j in a room, as shown on Figure 2. When a sound source S_1 generates a sound, TDoA measurements of the device pair D_i and D_j can be found using Eq.(2). Denote $\tau_{ij} \cdot v$ in Eq.(2) as $\hat{d}_{ij,1}$, which is the difference in distances from S_1 to D_i and D_j. According to the triangle inequality, $\hat{d}_{ij,1}$ sets a lower bound of the real inter-device distance d_{ij} such that $d_{ij} \geq |\hat{d}_{ij,1}|$. As the system detects more sounds from multiple sound sources, multiple distance differences establish a tighter lower bound of d_{ij} such that

$$d_{ij} \geq \max_{\forall m \in M}(|\hat{d}_{ij,m}|), \tag{11}$$

where $\hat{d}_{ij,m}$ is the difference in distances from the mth sound source to D_i and D_j, and M is the total number of sound sources that have generated ambient sounds.

As the number of sound sources increases, the maximum absolute value of all distance differences, i.e. $\max(|\hat{d}_{ij,m}|)$, will increase. Given an infinite number of uniformly distributed sound sources, $\max(|\hat{d}_{ij,m}|)$ will finally equal to the true inter-device distance d_{ij} if TDoA measurements have no errors. Thus $\max(|\hat{d}_{ij,m}|)$ can be regarded as a good estimate of d_{ij}. However, in a real indoor environment, TDoA measurements are corrupted by effects from reflections, obstacles and other ambient noises, and sound sources may not be uniformly distributed in the room. These factors make the value $\max(|\hat{d}_{ij,m}|)$ deviate significantly from the true value of d_{ij}.

To examine these effects during the estimation of inter-device distances d_{ij}, we build a geometry model in which sound sources are assumed uniformly distributed in the 2-D plane.

Two devices are located at $(-\frac{d_{ij}}{2}, 0)$ and $(\frac{d_{ij}}{2}, 0)$. We find that the probability density function (*pdf*) of the values of $\hat{d}_{ij}$ has three properties: 1) The *pdf* concentrates at values close to d_{ij} and $-d_{ij}$; 2) At values between $(-d_{ij}, d_{ij})$, the *pdf* resembles a uniform distribution; 3) At values outside $[-d_{ij}, d_{ij}]$, the *pdf* is 0.

To approximate this distribution, we define an empirical distribution such that: 1) In $(-d_{ij}, d_{ij})$, the values of $\hat{d}_{ij}$ are uniformly distributed; 2) At $\pm d_{ij}$, the probability equals ϵ times of that inside $(-d_{ij}, d_{ij})$; 3) Elsewhere, the probability is 0. Since the integration of the entire probability density function should be 1, we derive the probability distribution of $\hat{d}_{ij}$ as follows.

$$f_X(x) = \begin{cases} \frac{\epsilon}{2d_{ij}+2\epsilon} & x = \pm d_{ij} \\ \frac{1}{2d_{ij}+2\epsilon} & -d_{ij} < x < d_{ij} \\ 0 & \text{otherwise} \end{cases}. \tag{12}$$

For high SNR, TDoA values estimated by the GCC-PHAT algorithm are shown to be normally distributed with zero mean [15]. Therefore, we assume an additive Gaussian noise model $f_Y(y) \sim \mathcal{N}_{0,\sigma}$ for $\hat{d}_{ij}$ to take ambient noise into account. Together, the probability distribution of the noise-corrupted $\hat{d}_{ij}$ is the convolution of $f_X(x)$ and $f_Y(y)$, which is derived as

$$\begin{aligned} f_Z(z) &= \int_{-\infty}^{+\infty} f_X(z-y) f_Y(y) dy \\ &= \frac{\epsilon}{2d_{ij}+2\epsilon} \cdot (\mathcal{N}_{0,\sigma}(z-d_{ij}) + \mathcal{N}_{0,\sigma}(z+d_{ij})) \\ &\quad + \frac{1}{2}\left[\text{erf}\left(\frac{z-d_{ij}}{\sqrt{2}}\right) - \text{erf}\left(\frac{z+d_{ij}}{\sqrt{2}}\right)\right], \end{aligned} \tag{13}$$

where erf(.) denotes the error function.

Arrangement Detection for All Devices

Eq.(13) provides a parametric noise model of the distribution of $\hat{d}_{ij}$. It is determined by three parameters: 1) the actual inter-device distance d_{ij}, 2) the value of ϵ, and 3) the standard deviation of the Gaussian noise distribution σ. Therefore, we cannot simply take the maximum TDoA value as the true inter-device distance. However, given multiple TDoA measurements and thus multiple distance differences $\{\hat{d}_{ij,1}, \hat{d}_{ij,2}, ..., \hat{d}_{ij,m}\}$, d_{ij} can be estimated using a maximum likelihood estimator (MLE) that optimizes the following nonlinear optimization problem

$$\{d_{ij}, \epsilon, \sigma\}_{\text{MLE}} = \underset{d_{ij}, \epsilon, \sigma \in \mathcal{R}}{\arg\max} \prod_{m=1}^{M} f_Z(\hat{d}_{ij,m} | d_{ij}, \epsilon, \sigma). \tag{14}$$

Maximization of Eq.(14) can be solved using various nonlinear optimization techniques, such as the Levenberg-Marquardt algorithm [4]. The process of solving Eq.(14) is out of the scope of this work. Interested readers are referred to [4, 18].

One common problem with most nonlinear optimization techniques is that the optimization does not guarantee finding the global optimum unless given good initial starting points. Therefore, we create a heuristic algorithm that finds a close approximation of the real d_{ij} value. To begin with, each device pair D_i and D_j makes a histogram of all the historical $\hat{d}_{ij}$ values computed from all sound sources that have generated sound events. The algorithm consists of the following four steps, which are illustrated on Figure 3. First, each pair selects $a\%$ bins with the highest $\hat{d}_{ij}$ values to determine the location of the uniformly-distributed region. Second, the value of the largest absolute x-coordinate of all the selected bins is determined, which is denoted as p. The value of p can approximate the area where the Gaussian noise affects. Third, the value of the largest absolute x-coordinate of all the historical $\hat{d}_{ij}$ values is determined, which is denoted as q. The difference between p and q can approximate the standard deviation of the Gaussian noise. Finally, each pair uses $\frac{p+q}{2}$ as a starting point for the MLE process.

This heuristic algorithm cannot generate the actual value of d_{ij}, but since it's lightweight and provides a good estimate of d_{ij}, we use it to generate a starting point in the MLE process. Algorithm 3 outlines the process of generating the starting point.

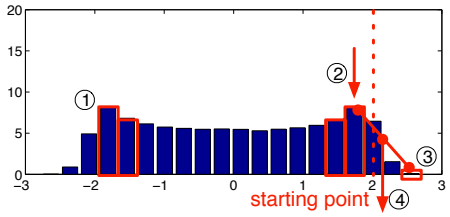

Figure 3. The four steps of the heuristic algorithm to determine the starting point for the MLE. Note that the histogram of the historical $\hat{d}_{ij}$ values shown is simulated based on the pair-wise TDoA noise model. In this example, the real inter-device distance d_{ij} is 2.

Algorithm 3 Starting Point Determination Algorithm

1: **for** each device pair D_i and D_j **do**
2: Make histogram of all historical $\hat{d}_{ij}$ values
3: Select $a\%$ bins that have the highest $\hat{d}_{ij}$ value (Step 1)
4: Let the value of the largest absolute x-coordinate of all the selected bins be p (Step 2)
5: Let the value of the largest absolute x-coordinate of all historical $\hat{d}_{ij}$ values of the pair be q (Step 3)
6: Determine the starting point as $\frac{p+q}{2}$ (Step 4)
7: **end for**

After the MLE process computes inter-device distances for all pairs, we compute the physical arrangement of all devices in the network using the multidimensional scaling (MDS) techniques [21]. Originally applied in the psychological field, the MDS is a set of statistical techniques for exploring similarities and dissimilarities in data. It takes a distance matrix that quantizes inter-point dissimilarity. In our case, the dissimilarity between points is given by the inter-device distances d_{ij}, which is estimated by $\hat{d}_{ij}$. The MDS creates a configuration of relative coordinates for the points, between which the Euclidean distances can approximately represent the original distance matrix. We use these output relative coordinates as the final estimates of the physical arrangement of the devices in the room.

SYSTEM IMPLEMENTATION AND EVALUATION

To evaluate the arrangement detection techniques, we present real experiments with a few sound source locations (fixed speakers to localize), and many locations (ambient sound localization). We implemented the PANDAA system using acoustic sensing nodes and deployed them in indoor environments. Each node is built upon an LPC-P2148 prototyping board, featuring an ARM7 60MHz CPU and 48KB memory, with an inexpensive microphone (Knowles MD9745APZ-F [1]) and a Bluetooth radio for diagnosis. Figure 4 (left) shows a PANDAA sensor node.

Experimental Traces and Setup

We conducted experiments in a $7.80 \times 5.96 \times 2.84\,\mathrm{m}^3$ meeting room in our building. The meeting room is a generic rectangular shape as well as it contains furniture including tables, chairs, a HDTV, a blackboard, and an electrical control desk. A picture of the room is shown on Figure 4 (right) and a floor plan of the room is shown on Figure 5. Eight PANDAA nodes were placed on a table at approximately

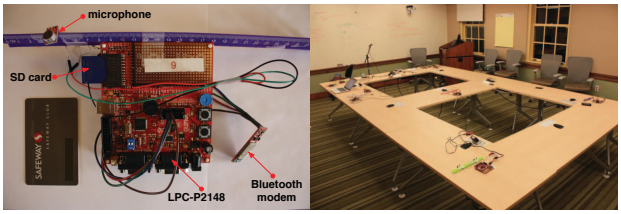

Figure 4. Left: A PANDAA acoustic sensing node placed next to a credit card. The microphone is fixed on a plastic ruler. Right: The meeting room where our controlled and full-scale experiments were done.

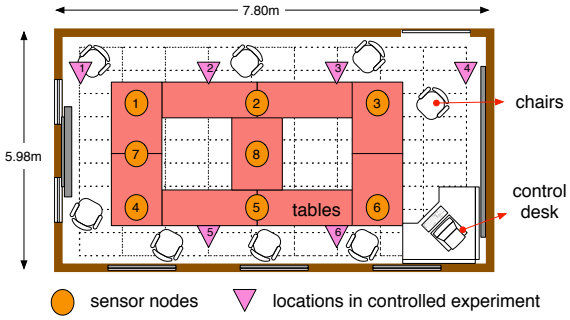

Figure 5. A floor plan of the meeting room. A total of 100 locations were marked as a grid (shown by dotted lines) in the meeting room and used as ground-truth locations of sound sources.

0.8m height as shown by the numbers in Figure 5. In order to make the system evaluation repeatable, a total of 100 locations were marked in the room, the intersection of the grid in Figure 5 shows the location of where ambient sounds were located.

To evaluate the system's ability of operating with different types of ambient sound, we selected cough, speech, and two genres of music as representative sound types common in indoor environments. All the sound was captured at 16KHz using a 10bit A/D converter on each of the eight PANDAA nodes. At the beginning of the experiments, a loud acoustic impulse was generated using a loudspeaker to synchronize all the nodes. During the experiments, one student was wandering inside the meeting room as well as a few chairs had been moved purposefully, generating time-varying ambient effects including noises and acoustic non-LoS.

Controlled Experiment

In order to test the capability of the system with a few fixed but repeatable sound locations (i.e. a few fixed speakers), we first performed a controlled experiment in the meeting room with six locations, as shown on Figure 5. At each location 160 human coughs were played using a loudspeaker. The pairwise TDoA values computed from the coughs were then used to analyze the TDoA errors. Figure 6 shows the distributions of errors, which approximately follow Gaussian distributions with zero mean. The standard deviations (std) for different device pairs changed slightly, with average std value of 1.27ms. Given the speed of sound, assumed to be 343m/s in the experiments, this std value translates to about 0.44m errors in inter-device distance estimation when sound sources are fixed but repeatable.

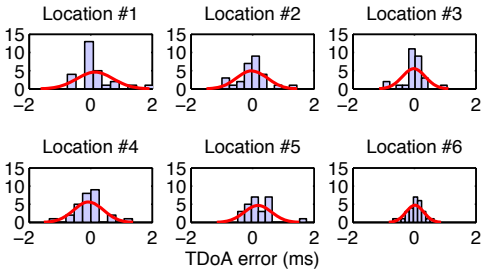

Figure 6. Distribution of pairwise TDoA errors for each sound source location with Gaussian fitting curves. Note that the number of devices is 8, so each histogram represents $\frac{8\times(8-1)}{2} = 28$ pairwise TDoA errors. The errors approximately follow Gaussian distributions.

Full-Scale Deployment

To evaluate performance of arrangement detection using unpredictable ambient sounds, we then tested in full-scale deployments using the 100 locations in the grid.

Performance of Sound Event Detection

To leverage a variety of ambient sounds, PANDAA detects impulsive sound events from different ambient sound signals recorded on each device. We selected coughs, speech and two different genres of music. The first genre was rock music (i.e. we chose Michael Jackson's "Billie Jean"), which features numerous clear rhythmic pulses that intuitively serve well to generate sufficient impulsive sounds; in contrast, the second genre was slower pop music (i.e. we chose Paul Simon and Garfunkel's "The Sound of Silence"), which was relatively smooth and melodic. In total, we tested all the four sound clips at each of the 100 locations in the full-scale experiments, which are summarized in Table 1.

Experimental results show that, the sound event detection algorithm is generalizable enough to handle the four different sound clips. Figure 7 shows the detection of impulsive sounds from one of the eight sensor nodes and Table 2 summarizes the frequency of detections and the durations of the detected events from all the nodes. Generally, every cough was considered by the nodes as a single sound event because of its distinct impulsive waveform. In addition, the "Billie Jean" song generated a large amount of sound events. This is explainable by considering its strong bass line. As a contrast, the "The Sound of Silence" song and human speeches generated considerably less sound events, but still achieved about one event per two seconds. On average, cough sounds generated 1 event/cough, and the other three sound clips generated 1 event/s. This frequency is sufficiently high to guarantee the system to acquire a large number of sound events quickly for arrangement detection. Using the detected sound events from each of the sound types, we computed TDoA values. Given the same sound source and device locations, the standard deviations of the TDoA measurements among different sound types were similar to those observed in the controlled experiments. Therefore, the following evaluations are based on the combination of different sound types.

Performance of TDoA Aggregation and Separation

As shown in the controlled experiment, raw TDoA values from the sound events contain considerable measurement

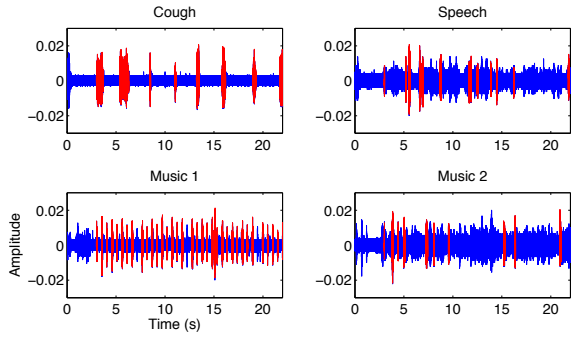

Figure 7. Sound events detected from the four sound clips. Blue: original sound clips. Red: detected sound events. While different sound types differ in time and spectrum properties, the event detection algorithm managed to detect and extract useful sound events from all the four tested sound clips.

Table 1. Ambient sounds used in the experiment

Type	Duration (s)	Note
Cough	32	12 coughs recorded by 6 individuals, including 2 males and 4 females
Human speech	21	A conversation between a male and a female
Music #1	21	"Billie Jean"
Music #2	21	"The Sound of Silence"

Table 2. Frequency and duration of detected sound events

Type	Freq. (events)	Avg. Duration (s)
Cough	1.10/cough	0.58
Human speech	0.67/s	0.24
Music #1 ("Billie Jean")	1.80/s	0.08
Music #2 ("The Sound of Silence")	0.55/s	0.17

errors due to indoor echoes and ambient noises. In PANDAA, the two-tier TDoA aggregation algorithm is aimed to suppress the time-domain uncorrelated noisy effects. In the upper-tier aggregation, accurately separating two consecutive sound events that come from different sound sources is important in improving TDoA performance. During the full-scale experiment, we randomly selected two locations from the grid to generate ambient sounds, forming a *sound-pair*. The ambient sounds were selected among the four sound clips. Then, multiple sound-pairs were considered by the system for aggregation or separation. Figure 8 shows the aggregation and separation results after the system considered 1000 randomly generated sound-pairs. The top graph indicates that there is a high positive correlation between the per-dimensional hybrid distances d_{ALL} of sound events and the real physical distances between the sound sources that have generated the events. The bottom graph shows that as the percentage of accurately aggregated sound events increases, the accuracy of sound events separation drops. This allows us to determine a value of the threshold d_{thr} such that the number of accurate aggregation can be maximized while maintaining a low false separation rate. During the experiment, we empirically set d_{thr} to 1.80. This led to 92.5% aggregation accuracy and 93.9% separation accuracy. After aggregating consecutive sound events, the standard deviation of TDoA errors are reduced by about 36%. Table 3 summa-

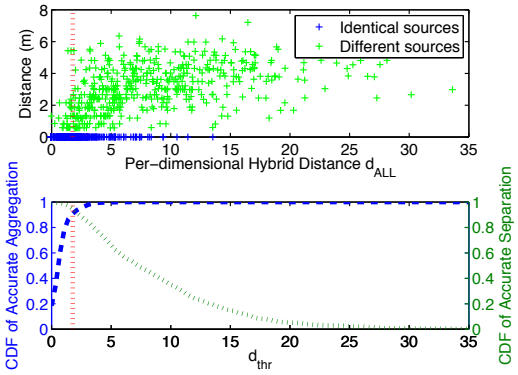

Figure 8. Upper-tier TDoA aggregation. Top: Per-dimensional hybrid distances between consecutive sound events vs. the physical distances between sound sources that generate the events. Note that if the two sound sources are identical, their physical distance (i.e. the y-coordinate) is 0. Bottom: Value of d_{thr} affects the accuracy of aggregating TDoAs from identical sources as well as separating TDoAs from different sources.

rizes the results comparing the standard deviation of TDoA errors between using and without using the TDoA aggregations.

Table 3. Comparison of pairwise TDoA estimation performance between using and without using the two-tier aggregation

Std. of TDoA errors (ms)	No Aggregation	With Aggregation
Cough	1.86	1.16
Human speech	2.14	1.02
Music #1	1.95	1.40
Music #2	2.18	1.64

Pairwise TDoA Error Rejection Performance

During the experiment, sound sources were randomly located at the 100 grid points shown in the room. After a sound source was selected, its ambient sound type was chosen from the four sound clips and a sound duration was set as a random value between 0s to the maximum clip durations shown in Table 1. The two-tier aggregation was enabled to improve accuracy of TDoA measurements. Figure 9 shows the performance of arrangement detection. Four approaches to compensating TDoA errors were compared:

1. Max: using the maximum historical pairwise TDoA measurements as estimates of inter-device distances d_{ij}. This approach was used as a baseline.
2. Heuristic: using the starting points computed in Algorithm 3 as estimates of d_{ij}.
3. MLE: using the starting points to initialize the MLE and use the MLE to estimate d_{ij}.
4. MLE ($\epsilon = 0$): assuming $\hat{d}_{ij}$ is uniformly distributed.

The figure shows that, as the number of sound sources increases, the errors of coordinate estimations drop quickly. For approaches except the Max baseline, the estimation performance becomes stable after about 10 sound sources having generated ambient sounds. After the estimation becomes stable, the baseline approach achieves the highest estimation

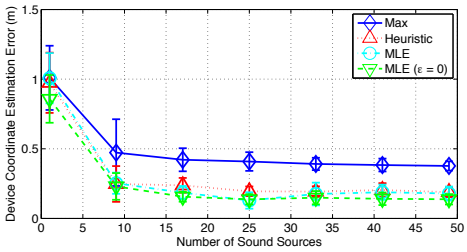

Figure 9. Performance of arrangement detection when using different methods to determine the inter-device distance.

error. Since pairwise TDoA errors are compensated in the second and third approaches, both of these two approaches reduce errors by at least 35%.

During the experiments we observed that the performance of MLE varied considerably between different evaluation runs. This was due to the fact that the MLE method tended to overfit the multiple TDoA measurements given that the number of sound sources was small, which caused the value of d_{ij} to oscillate significantly. Many techniques can be applied to mitigate over-fitting in the MLE, such as the penalty-based approaches or reducing the number of parameters [6]. After testing different approaches, we chose an efficient approach that reduced the number of parameters in MLE by setting ϵ to 0. This change actually makes the values of $\hat{d}_{ij}$ to be uniformly distributed between $[-d_{ij}, d_{ij}]$. As shown on Figure 9, the fourth approach "MLE ($\epsilon = 0$)" leads to the highest accuracy of coordinate estimations, which is 0.17m. This performance is about 2.5X better than the baseline Max approach.

Performance with Different Numbers of Sensor Nodes

We evaluate the performance limit of PANDAA when the number of nodes is limited. Since at least three nodes are needed to uniquely define a 2-D plane, we start to evaluate the system by using 4, 6, and 8 nodes. Figure 10 shows the accuracy improvement as the system evolves under different node numbers. In the 4-node case, the convergence of coordinate estimation accuracy is slower than those under 6- and 8-node cases, which requires about 20 sound sources. This is because under fewer node numbers, the system requires more sound sources to make accurate estimation of inter-device distances. In contrast, we do not observe significant difference in the ultimate accuracy among the three cases. This is because as the number of sound sources increases, eventually the system will always have high probability of accurately estimating inter-device distances for every device pair. This result shows that the performance of PANDAA is robust against changes in the number of devices and is mostly determined by estimation accuracy of inter-device distances.

Performance Under Skewed Source Distributions

In real scenarios, different furniture layouts, room shapes and resident life patterns may cause sound sources to not be uniformly distributed (i.e. people might only walk in one part of the room). We would like to evaluate how the performance of arrangement detection changes given skewed distributions of source locations.

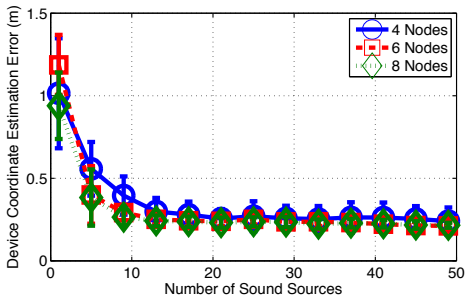

Figure 10. Performance comparison when using different numbers of sensor nodes.

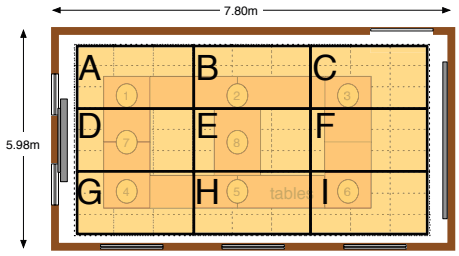

Figure 11. The location grid is divided into nine blocks to evaluate effects from skewed source distributions.

Therefore, we first divide the grid into nine blocks, as shown on Figure 11, and group them into symmetrical groups. We define four Side regions (for example, block A, B, and C form the Up-Side region), four Corner regions (for example, block A, B, D, and E form the Up-Left Corner region), and Center regions (containing block B, D, E, F, and H). Then, we select sound source locations *only* from within one of these regions to evaluate the system. Table 4 summarizes the performance under different cases.

To consider the convergence process and wait for the performance of arrangement detection to become stable, the estimation accuracy is determined after 20 sound sources have generated sounds. For all cases, skewed distribution of sound sources had higher errors than that of the uniformly distribution case described earlier. In particular, sound sources that are limited to only one side of the room (Side regions) have higher errors due to most limited diversity of sound source locations. However, we observe that PANDAA still achieved better than 0.4m estimation accuracy.

Table 4. Accuracy of arrangement detection under skewed source distributions (Accuracy is determined after 20 sound sources as the performance tends to be stable)

Cases	Avg. Error (m)	Std. Error (m)
Side Regions	0.39	0.17
Corner Regions	0.21	0.08
Center Regions	0.27	0.06

Using PANDAA to Determine Source Locations

As shown in previous evaluation, under uniformly distributed sound sources, PANDAA achieves up to 0.17m in accuracy. We would like to investigate how the performance of arrangement detection affects that of sound source localization. Figure 12 shows that, as the number of sensor nodes

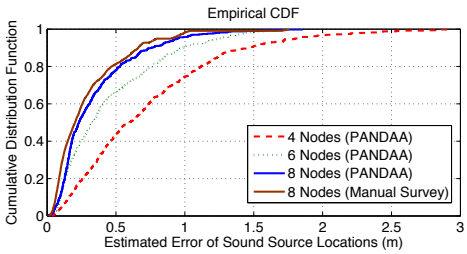

Figure 12. Sound source localization performance when using the PANDAA system.

increases in the system, the accuracy of source localization improves significantly. Using full sound source distributions, with 8 sensor nodes, PANDAA can localize a sound source to within 0.5m 80% of the time, and within 0.9m 95% of the time. This performance is comparable to previous work that localizes sound sources based on manually surveyed sensor node locations [10, 25]. Meanwhile, we compare PANDAA's performance with that of manually surveyed device locations. Under the 8-node case, the performance of manually surveyed locations is only about 10.5% better than PANDAA.

RELATED WORK AND DISCUSSION

There are many techniques that have been previously explored for indoor localization. The Global Positioning System (GPS) [3] is only available outdoors and provides limited accuracy. In indoor settings, researchers have developed RF radio-based positioning techniques like Radio Signal Strength (RSS) ranging, RSS fingerprinting, Angle-of-Arrival, RF Time Difference of Arrivals (TDoA), and Time of Flight (ToF) [16]. However, localization is not good enough for sub-meter accuracy required in many applications. Some work using ultrasonic sensing provides higher accuracy [12, 19], but is limited due to the use of specialized sensors, which are not ubiquitously available on electronic devices, and often require specialized configuration or setup to produce results.

Other work has shown audible sound can be a good resource for localizing devices. However, most of the previous work uses on-device loudspeakers to actively emit known finite-length repeatable signals, such as chirps or peaks [9, 17, 20]. This approach can be intrusive and sometimes annoying if the systems require frequent location updates of their devices.

Ambient sounds have also been used in outdoor environments to passively determine locations of devices [14]. While they report 0.58m average microphone localization error in open outdoor scenarios, their simulations show that the system generates large errors (>2m) when sound sources are close to the microphones indoors. In contrast, PANDAA deals with indoor ambient effects by selectively using impulsive sounds to compute TDoA, and leveraging multi-node collaboration to improve measurement accuracy. Scott and Dragovic also describe an audio location sensing technique that uses impulsive sounds [22], however it is not clear how the system deals with indoor acoustic effects, such as echoes and non-LoS.

TDoA is a common technique to self-localize devices in microphone arrays [2]. Prior work has discussed this technique using both close-formed [13] or iterative methods [26] theoretically. For distributed sensor networks, Raykar et al. present an approximate solution combined with nonlinear optimization [20]. However, they make an additional assumption that some loudspeakers are co-located with microphones. In contrast, PANDAA assumes no a priori knowledge of ambient sound types or locations of sound sources, and we present generalizable impulsive sound event detection and TDoA aggregation algorithms to leverage ambient sound in arrangement detection. Other TDoA approaches for sound source localization have also been explored. Guo and Hazas make comprehensive comparisons between TDoA estimation techniques in terms of sound type, accuracy, power and bandwidth requirements [10], but assume fixed microphones with known locations.

This paper focuses on the design and evaluation of PANDAA, thus implementation details are not fully addressed due to the page limit and are presented elsewhere [24].

CONCLUSION

In this paper, we presented PANDAA, a novel autonomous physical arrangement detection technique that determines device locations using ambient sounds generated in indoor environments. Using ambient sounds commonly existing in the office and home environments (i.e. coughs, human speech, and music), the system is able to accurately measure device location. To show the performance of PANDAA in real deployments, we deployed our system in a meeting room environment under different source distributions, and with different sound types. Using only TDoA as a bound for its successive measurements, PANDAA was able to achieve an average of 0.17m device location accuracy. In addition, based on this estimation, PANDAA is able to localize sound sources to within 0.5m 80% of the time, well within comparable error level to previous work that is based on manual surveys.

By autonomously resolving the spatial relative arrangement using trigonometric bounds and successive approximation, PANDAA is less intrusive and more accurate than existing techniques. We propose our system can be used to solve the device localization problem for many emerging indoor ubiquitous computing applications. Furthermore, our automatic localization of ubiquitous computing devices has the potential to greatly facilitate device deployments in future smart home environments.

REFERENCES

1. Knowles microphone, www.knowles.com/search/prods_pdf/MD9745APZ-F.pdf, 2011.
2. P. Aarabi. Self-Localizing Dynamic Microphone Arrays. *IEEE Transactions on Systems, Man and Cybernetics*, 32(4):474–484, Nov. 2002.
3. A. Ali, S. Asgari, T. Collier, M. Allen, L. Girod, R. Hudson, K. Yao, C. Taylor, and D. Blumstein. An Empirical Study of Collaborative Acoustic Source Localization. In *Proc. of the 6th International Conference on Information Processing in Sensor Networks*, volume 57, pages 415–436. Springer, Nov. 2007.
4. D. P. Bertsekas. *Nonlinear Programming*. Athena Scientific, 1999.
5. X. Bian, G. Abowd, and J. Rehg. Using sound source localization to monitor and infer activities in the Home. In *Proc. of the 3rd International Conference on Pervasive Computing*, Munich, Germany, 2005. Springer.
6. C. Bishop. *Pattern Recognition and Machine Learning*, volume 4. Springer New York, 2006.
7. M. Brandstein and H. Silverman. A Practical Methodology for Speech Source Localization with Microphone Arrays. *Computer Speech & Language*, 11(2):91–126, Apr. 1997.
8. J. DiBiase. *A High-Accuracy , Low-Latency Technique for Talker Localization in Reverberant Environments Using Microphone Arrays*. PhD thesis, 2000.
9. L. Girod, V. Bychkovskiy, J. Elson, and D. Estrin. Locating Tiny Sensors in Time and Space: A Case Study. In *Proc. of the IEEE International Conference on Computer Design: VLSI in Computers and Processors 2002*, pages 214–219. IEEE, 2002.
10. Y. Guo and M. Hazas. Localising Speech, Footsteps and Other Sounds Using Resource-Constrained Devices. In *Proc. of the 10th International Conference on Information Processing in Sensor Networks*, number 2, pages 330 – 341, Chicago, Illinois, USA, 2011. IEEE.
11. A. Harma, M. McKinney, and J. Skowronek. Automatic Surveillance of the Acoustic Activity in Our Living Environment. In *Proc. of the IEEE International Conference on Multimedia and Expo 2005*, volume 1, pages 1–4, Zurich, Switzerland, 2005. IEEE.
12. M. Hazas, C. Kray, H. Gellersen, H. Agbota, G. Kortuem, and A. Krohn. A Relative Positioning System for Co-located Mobile Devices. In *Proc. of the 3rd International Conference on Mobile Systems, Applications, and Services*, pages 177–190, New York, New York, USA, 2005. ACM.
13. K. Ho. An Approximately Efficient TDOA Localization Algorithm in Closed-Form for Locating Multiple Disjoint Sources with Erroneous Sensor Positions. *IEEE Transactions on Signal Processing*, 57(12):4598–4615, Dec. 2009.
14. T. Janson, C. Schindelhauer, and J. Wendeberg. Self-Localization Application for iPhone Using Only Ambient Sound Signals. In *Proc. of the International Conference on Indoor Positioning and Indoor Navigation 2010*, number September, pages 1–10. IEEE, 2010.
15. C. Knapp and G. Carter. The Generalized Correlation Method for Estimation of Time Delay. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 24(4):320–327, 1976.
16. H. Liu, H. Darabi, P. Banerjee, and J. Liu. Survey of Wireless Indoor Positioning Techniques and Systems. *IEEE Transactions on Systems Man and Cybernetics*, 37(6):1067–1080, 2007.
17. C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan. BeepBeep: A High Accuracy Acoustic Ranging System Using COTS Mobile Devices. In *Proc. the 5th International Conference on Embedded Networked Sensor Systems*, pages 1–14, Sydney, Australia, 2007. ACM.
18. W. H. Press, B. P. Flannery, S. A. Teukolsky, and W. T. Vetterling. *Numerical Recipes in C: The Art of Scientific Computing*. Cambridge University Press, 1992.
19. N. Priyantha, A. Chakraborty, and H. Balakrishnan. The Cricket Location-Support System. In *Proc. of the 6th International Conference on Mobile Computing and Networking*, volume 2000, pages 32–43, Boston, Massachusetts, USA, 2000. ACM.
20. V. Raykar, I. Kozintsev, and R. Lienhart. Position Calibration of Microphones and Loudspeakers in Distributed Computing Platforms. *IEEE Transactions on Speech and Audio Processing*, 13(1):70–83, Jan. 2005.
21. S. Schiffman, M. L. Reynolds, and F. W. Young. *Introduction to Multidimensional Scaling: Theory, Methods, and Applications*. Emerald Group Publishing Limited, 1981.
22. J. Scott and B. Dragovic. Audio Location: Accurate Low-Cost Location Sensing. In *Proc. of the 3rd International Conference on Pervasive Computing*, pages 1–18, Munich, Germany, 2005. Springer.
23. M. Silverman. A Robust Method for Speech Signal Time-Delay Estimation in Reverberant Rooms. In *Proc. of the IEEE International Conference on Acoustics, Speech and Signal Processing 1997*, pages 375–378. IEEE, 1997.
24. Z. Sun, A. Purohit, P. De Wagter, I. Brinster, C. Hamm, and P. Zhang. PANDAA: A Physical Arrangement Detection Technique for Networked Devices through Ambient-Sound Awareness. In *Proc. of the ACM SIGCOMM 2011*, pages 1–2, Toronto, Canada, 2011.
25. Z. Sun, A. Purohit, K. Yang, N. Pattan, D. Siewiorek, I. Lane, and P. Zhang. CoughLoc : Location-Aware Indoor Acoustic Sensing for Non-Intrusive Cough Detection. In *Proc. of the International Workshop on Emerging Mobile Sensing Technologies, Systems, and Applications 2011*, pages 1–6, San Francisco, California, USA, 2011.
26. D. J. Torrieri. Statistical Theory of Passive Location Systems. *IEEE Transactions on Aerospace and Electronic Systems*, pages 183– 198, July 1984.