Non-Field-of-View Acoustic Target Estimation in Complex Indoor Environment

Kuya Takami, Tomonari Furukawa, Makoto Kumon and Gamini Dissanayake

Abstract This paper presents a new approach which acoustically localizes a mobile target outside the Field-of-View (FOV), or the Non-Field-of-View (NFOV), of an optical sensor, and its implementation to complex indoor environments. In this approach, microphones are fixed sparsely in the indoor environment of concern. In a prior process, the Interaural Level Difference (ILD) of observations acquired by each set of two microphones is derived for different sound target positions and stored as an acoustic cue. When a new sound is observed in the environment, a joint acoustic observation likelihood is derived by fusing likelihoods computed from the correlation of the ILD of the new observation to the stored acoustic cues. The location of the NFOV target is finally estimated within the recursive Bayesian estimation framework. After the experimental parametric studies, the potential of the proposed approach for practical implementation has been demonstrated by the successful tracking of an elderly person needing health care service in a home environment.

1 Introduction

Target localization and tracking, or mobile target estimation, in indoor environments has been a research challenge over several decades due to the existence of a variety of applications in addition to the significance and the difficulty of each application. It is significant in applications such as home security, home health care and urban search-and-rescue but its usefulness is limited by the complexity of indoor structures [13, 7]. Complex indoor structures make estimation problems challenging as they can introduce large unobservable regions when an optical sensor such as a camera is deployed. This is because optical sensors' FOV is determined by

Kuya Takami and Tomonari Furukawa
Department of Mechanical Engineering, Virginia Tech,Blacksburg, VA, USA
e-mail: {kuya,furukawa}@vt.edu

Makoto Kumon
Department of Mechanical System Engineering, Kumamoto University, Kumamoto, Japan
e-mail: kumon@gpo.kumamoto-u.ac.jp

Gamini Dissanayake
Center for Autonomous Systems, University of Technology, Sydney, NSW, Australia
e-mail: gamini.dissanayake@uts.edu.au

the Line-of-Sight (LOS) and range of the optical sensor, which could be small in highly constrained environments. In addition, there are environments such as personal homes where privacy concerns do not allow for the use of cameras. These limitations on optical sensors give rise to need for NFOV mobile target estimation.

Recent work for NFOV mobile target estimation has been tackled in three different ways. The first approach deploys target mounted radio-frequency (RF) transmitters and fixed receivers in the environment. In one arrangement, RF receivers form a wireless sensor network (WSN), and numerical techniques are used to localize a NFOV target by processing information of received signals such as signal intensity [3, 6]. An improved arrangement with minimal infrastructure uses "fingerprints" [1, 10]. There is a unique fingerprint at each location in a static environment. A target can thus be localized by feature-matching the fingerprints. Whilst this arrangement can achieve higher accuracy, the critical problem inherent in the RF based approach is its applicability only to near-NFOV target estimation [13, 15].

In the second approach, acoustic sensors are used for target estimation. Since sound signals are reflected by structures, it is possible to localize a NFOV target unlike the RF based approach provided that the sound signals contain information on the target location. The most common approach utilizes the Time-of-Arrival (TOA)/Time-Difference-of-Arrival (TDOA) information of acoustic signals [2, 18, 11]. The existing acoustic techniques, however, have not achieved true NFOV target estimation to the best of our knowledge. The majority of sound localization challenges have been focused on the direction of sound rather than its position due to complexity of sound wave propagation [17, 16].

The final approach enhances NFOV target estimation by including a sensor with a limited FOV, such as an optical sensor, by applying a numerical technique. Mauler [12] stated the NFOV estimation problem mathematically, and Furukawa, *et al.* [4, 5] developed a generalized numerical solution. In this technique, the event of "no detection" is converted into an observation likelihood and utilized to positively update probabilistic belief on the target. This belief is dynamically maintained by the recursive Bayesian estimation (RBE). The technique, however, has been found to fail in target estimation unless the target is re-discovered within a short period after being lost. Kumon, *et al.* [9], incorporated an acoustic sensor to maintain belief with no optical detection more reliably. Nevertheless, the technique performed poorly unless the target re-entered the optical FOV since the acoustic sensing is only conducted in an assistive capacity.

This paper presents a new acoustic approach to estimate a NFOV mobile target, and its application and implementation to complex indoor environments. In the approach, microphones are sparsely installed in an indoor environment. In a prior process to the estimation, the ILD of observations acquired by combination of stereo microphone pairs is derived for different target positions and stored as the "fingerprints", or acoustic cues. This *a priori* data collection process is accelerated by a speaker localization device. With the acquisition of a new sound from the target, an acoustic observation likelihood is computed for dominant pair of microphones by quantifying the correlation of the ILD of the new observation to the stored ILDs. The joint likelihood is then created by fusing the acoustic observation likelihoods,

and the NFOV target is estimated by recursively updating the belief within the RBE framework using the joint likelihood.

2 Recursive Bayesian Estimation

Consider the motion of a target t, which is discretely given by

$$\mathbf{x}_{k+1}^t = \mathbf{f}^t\left(\mathbf{x}_k^t, \mathbf{u}_k^t, \mathbf{w}_k^t\right) \tag{1}$$

where $\mathbf{x}_k^t \in \mathcal{X}^t$ is the target state at time step k, $\mathbf{u}_k^t \in \mathcal{U}^t$ is the set of control inputs, and $\mathbf{w}_k^t \in \mathcal{W}^t$ is the "system noise". For simplicity, the target state describes the two-dimensional position.

FOV and NFOV are defined by physical properties of a camera s_c where the global state of the optical sensor is assumed to be known as $\tilde{\mathbf{x}}^s \in \mathcal{X}^s$. Note that $\tilde{()}$ is an instance of $()$. The FOV of the optical sensor can be expressed by the probability of detecting the target $P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^{s_c}\right)$ as ${}^{s_c}\mathcal{X}_o^t = \{\mathbf{x}_k^t|0 < P_d\left(\mathbf{x}_k^t|\tilde{\mathbf{x}}^{s_c}\right) \leq 1\}$. Accordingly, the target position observed from the optical sensor, ${}^{s_c}\mathbf{z}_k^t \in \mathcal{X}^t$, is given by

$${}^{s_c}\mathbf{z}_k^t = \begin{cases} {}^{s_c}\mathbf{h}^t\left(\mathbf{x}_k^t, \tilde{\mathbf{x}}^s, {}^{s_c}\mathbf{v}_k^t\right), & \text{if } \mathbf{x}_k^t \in {}^{s_c}\mathcal{X}_o^t \\ \varnothing, & \text{otherwise} \end{cases} \tag{2}$$

where ${}^{s_c}\mathbf{h}^t$ is the optical sensor model, ${}^{s_c}\mathbf{v}_k^t$ is the observation noise, and $\varnothing$ represents an "empty element", indicating that the optical observation contains no information on the target or that the target is unobservable when it is not within the observable region. The acoustic sensor can, on the other hand, observe a target on the Non-Line-of-Sight (NLOS) or even in the NFOV with limited accuracy due to the complex behavior of sound signals including reflection, refraction and diffraction. Because of its broad range, the observation region of the acoustic sensor could be considered unlimited when compared to that of the optical sensor. The acoustic sensor model ${}^{s_a}\mathbf{h}^t$ can be then constructed without defining an observable region unlike the optical sensor model:

$${}^{s_a}\mathbf{z}_k^t = {}^{s_a}\mathbf{h}^t\left(\mathbf{x}_k^t, \tilde{\mathbf{x}}^s, {}^{s_a}\mathbf{v}_k^t\right) \tag{3}$$

The RBE updates belief on a dynamical system, given by a probability density, in both time and observation. Let a sequence of observations of a moving target t by a stationary sensor system s from time step 1 to time step k be ${}^s\tilde{\mathbf{z}}_{1:k}^t \equiv \{{}^s\tilde{\mathbf{z}}_\kappa^t|\forall\kappa \in \{1, ..., k\}\}$. Given the initial belief $p\left(\mathbf{x}_0^t\right)$, the sensor platform state $\tilde{\mathbf{x}}^s$ and a sequence of observations ${}^s\tilde{\mathbf{z}}_{1:k}^t$, the belief on the target at any time step k, $p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t, \tilde{\mathbf{x}}^s\right)$ can be estimated recursively through the two stage equations. The prediction may be expressed as

$$p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right) = \int_{\mathcal{X}^t} p\left(\mathbf{x}_k^t|\mathbf{x}_{k-1}^t\right) p\left(\mathbf{x}_{k-1}^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t, \tilde{\mathbf{x}}^s\right) d\mathbf{x}_{k-1}^t, \tag{4}$$

whereas the correction takes the form

$$p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k}^t,\tilde{\mathbf{x}}^s\right) = \frac{l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}^s\right)p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right)}{\int_{\mathcal{X}^t} l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}^s\right)p\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_{1:k-1}^t,\tilde{\mathbf{x}}^s\right)d\mathbf{x}_{k-1}^t}, \tag{5}$$

where $l\left(\mathbf{x}_k^t|{}^s\tilde{\mathbf{z}}_k^t,\tilde{\mathbf{x}}^s\right)$ represents the likelihood of $\mathbf{x}_k^t$ given ${}^s\tilde{\mathbf{z}}_k^t$ and $\tilde{\mathbf{x}}^s$, which is a probabilistic version of the sensor model; i.e., Equation (2) if the sensor is optical. It is to be noted that the likelihood does not need to be a probability density since the normalization in Equation (5) makes the output belief a probability density regardless of the formulation of the likelihood.

3 NFOV Acoustic Target Estimation

3.1 Indoor Installation

Figure 1 shows a schematic for the hardware installation necessary for the proposed acoustic target estimation approach. As shown in the figure, microphones are placed with some distance in the indoor environment. This is a complex environment where optical sensors could not be used effectively as a large number of optical sensors would need to be placed to cover the entire space. Microphones, on the other hand, can collect information on the NFOV. A much lower number of inexpensive sensors need to be installed for this reason, making the installation efficient in both time and cost.

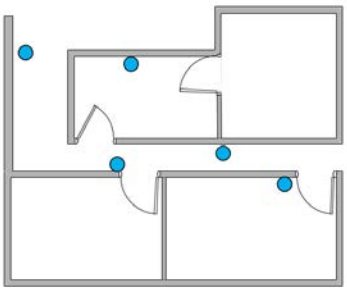


Fig. 1 Schematic of hardware installation for proposed approach where circles indicate microphones

3.2 Modeling of Acoustic Observation Likelihood

In accordance with the preliminary investigations of the authors [8], the theoretical approach proposed in this paper constructs acoustic cues of the target in the environment of concern *a priori* to create an acoustic observation likelihood. The assumption of two-dimensional (2D) space and a use of a data collection device in the proposed method reduce the time consumed by *a priori* data collection. First, the three-dimensional (3D) complex environment can be simplified by assuming the omni-directional sound source belongs in the 2D planar domain depicted in Figure 2. This assumption is realized by placing a sound source at a foot level which

generally kept at constant height throughout movement of a human. Second, *a priori* sound data is collected automatically using a speaker with range finders, which measures the distance to the walls to locate the speaker and emits white noise when the data collection button is pressed.

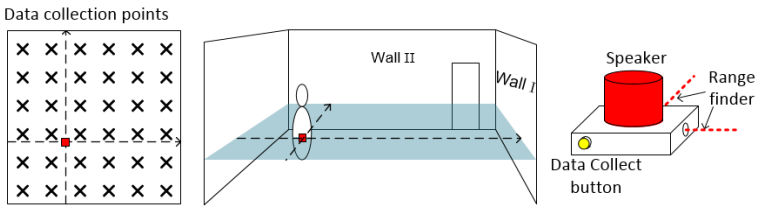


Fig. 2 Data collection and localization

Having the data collected into the ILD database in the prior process, fig. 3 shows a schematic diagram of the main process of the proposed approach. Given the target sound, The acoustic observation likelihood is created for each microphone pair by correlating the observation with ILD vectors in the database. The collection of observation likelihood finally yields a joint acoustic observation likelihood. This fusion process only considers few dominant microphone pairs above the signal-to-noise ratio (SNR) threshold for scalability of the system.

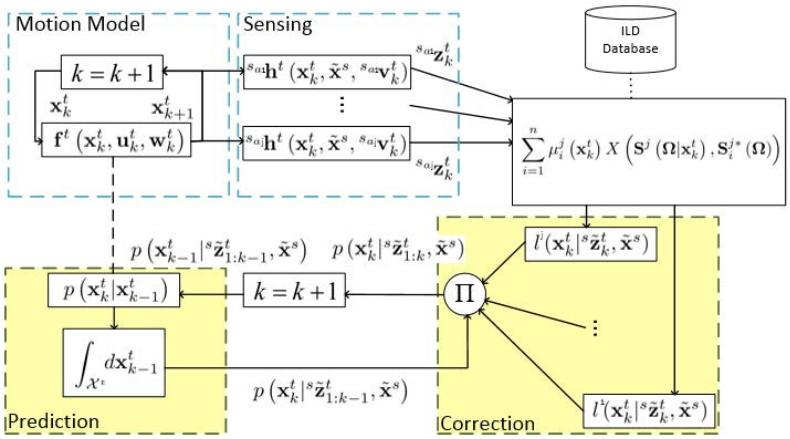


Fig. 3 Schematic diagram of proposed approach within the RBE framework

Mathematically, let the estimation of the *a priori* i-th data collection position be $(\tilde{\mathbf{x}}_k^t)_i$. When a target sound is observed by j_{m}-th microphone at $\tilde{\mathbf{x}}_k^s$, the sound is considered "detected" if the SNR of the microphone is greater than the SNR threshold:

$$s_{j_{\mathrm{m}}}^{\mathrm{S}} \equiv \frac{s_{j_{\mathrm{m}}}\left(\omega|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)}{s_{j_{\mathrm{m}}}\left(\omega\right)_{ambient}} > \delta_{\mathrm{S}} \tag{6}$$

where ω is the sound frequency. Stereo microphone pairs increase with combination of form $\binom{n}{r} = \frac{n!}{r!(n-r)!}$ by choosing stereo pair $r = 2$ from n possible microphones. Figure 4 shows the detectable region of red and yellow microphone as the j-th mi-

crophone pair $\{j_1, j_2\}$. When the target is located within union of those regions, the ILD of the microphone pair is constructed:

$$\mathbf{x}^t \in {}^{s_a}\mathcal{X}_d^t(\gamma, \delta_{\mathrm{S}}) = {}^{s_{aj_1}}\mathcal{X}_d^t(\gamma_{j_1}, \delta_{\mathrm{S}}) \cap {}^{s_{aj_2}}\mathcal{X}_d^t(\gamma_{j_2}, \delta_{\mathrm{S}}). \tag{7}$$

where γ is acoustic and environmental characteristics. It is reasonable to sort and

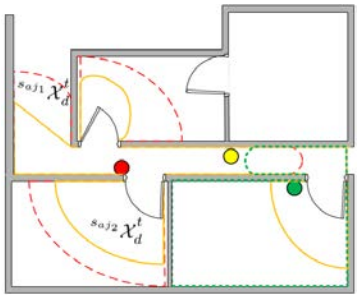


Fig. 4 Detectable region indicated by lines for each microphone location

choose the microphones with largest s^{S} values. The maximum microphone pair is set to be $j_{\max}$. Following the above selection process, the ILD of the j-th microphone pair $\{j_1, j_2\}$ for the i-th position $(\tilde{\mathbf{x}}_k^t)_i$, $\Delta S_i^j(\omega)$, is then given by

$$\Delta S_i^j(\omega) = 20\log\left|s_{j_1}\left(\omega|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right| - 20\log\left|s_{j_2}\left(\omega|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|. \tag{8}$$

If the ILD is sampled at N frequencies $\boldsymbol{\Omega} = [\omega_1, ..., \omega_N]^\top$, the ILD vector can be described as

$$\mathbf{S}_i^j(\boldsymbol{\Omega}) = \left[a_1^j \Delta S_i^j(\omega_1), ..., a_N^j \Delta S_i^j(\omega_N)\right]^\top, \tag{9}$$

where

$$a_i^j = \left\langle \min\{\left|s_{j1}\left(\omega_N|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|, \left|s_{j2}\left(\omega_N|\left(\tilde{\mathbf{x}}_k^t\right)_i\right)\right|\} - \epsilon \right\rangle. \tag{10}$$

In the equation, $\langle\cdot\rangle$ is Macaulay brackets, and $\min\{\cdot,\cdot\}$ returns the smaller value of the two entities. The acoustic observation likelihood modeling results in the ILD vectors for n target positions, i.e., $\mathbf{S}_i^{j*}(\boldsymbol{\Omega}), \forall i \in \{1, ..., n\}$. They are essentially the acoustic cues to be prepared in advance and used to create the acoustic observation likelihood. The selection of microphone pairs $\mathbf{S}_i^{j*}(\boldsymbol{\Omega})\, \forall j \in \{1, ..., j_{\max}\}$ must satisfy the conditions $s_j^{\mathrm{S}} > \delta_{\mathrm{S}}$.

Given the ILD vector $\mathbf{S}^j(\boldsymbol{\Omega}|\mathbf{x}_k^t)$ created from ${}^s\tilde{\mathbf{z}}_k^t$ with the unknown target position $\mathbf{x}_k^t$, the proposed technique quantifies its degree of correlation to the i-th ILD vector as

$$X\left(\mathbf{S}^j\left(\boldsymbol{\Omega}|\mathbf{x}_k^t\right), \mathbf{S}_i^{j*}(\boldsymbol{\Omega})\right) = \frac{1}{2}\left\{\frac{\mathbf{S}^j(\boldsymbol{\Omega}|\mathbf{x}_k^t)^\top \mathbf{S}_i^{j*}(\boldsymbol{\Omega})}{|\mathbf{S}^j(\boldsymbol{\Omega}|\mathbf{x}_k^t)|\left|\mathbf{S}_i^{j*}(\boldsymbol{\Omega})\right|} + 1\right\}. \tag{11}$$

where $0 \leq X(\cdot) \leq 1$. The acoustic observation likelihood with the particular $\mathbf{S}_m(\mathbf{\Omega}|\mathbf{x}_k^t)$ can be finally calculated as

$$l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \sum_{i=1}^{n} \mu_i^j\left(\mathbf{x}_k^t\right) X\left(\mathbf{S}^j\left(\mathbf{\Omega}|\mathbf{x}_k^t\right), \mathbf{S}_i^{j*}(\mathbf{\Omega})\right), \tag{12}$$

where $\mu_i^j(\mathbf{x}_k^t)$ is a basis function developed by adjacent measurements. One of the suited basis function is a T-spline basis function where $\mu_i^j(\mathbf{x}_k^t)$ in a T-mesh in parameter space (s,t) can be represented as

$$\mu_{im}(s,t) = g(s)g(t) \tag{13}$$

where $g(s)$, and $g(t)$ are the cubic B-spline basis functions. Further detailed formulations are found in [14]. Similarly to $X(\cdot)$, $l_j^a(\cdot)$ is also bounded as $0 \leq l_m^a(\cdot) \leq 1$ due to the use of the shape function.

Finally, the joint likelihood is derived by the canonical data fusion formula:

$$l^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right) = \prod_j l_j^a\left(\mathbf{x}_k^t|^s\tilde{\mathbf{z}}_k^t, \tilde{\mathbf{x}}_k^s\right). \tag{14}$$

4 Numerical and Experimental Analysis

The efficacy of the proposed approach was examined experimentally in two steps. The first step was aimed at studying the capabilities and limitations of the proposed acoustic sensing technique by parametrically changing the complexity of the test environment. This was accomplished with an experimental system consisting of a speaker array and a movable/replaceable wall developed specifically for this study. After verifying the feasibility of the acoustic sensing technique for NLOS target localization, the applicability of the proposed approach in a practical indoor scenario was investigated. The investigation looked into not only the performance of the proposed approach but also compared it to a conventional approach.

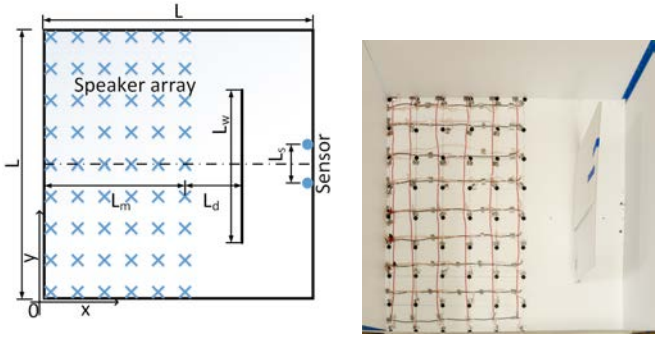
4.1 Acoustic Observation of NLOS Target

Figure 5(a) shows the design of the experimental system that changed the complexity of the environment for the evaluation of the proposed approach. The number of microphones was fixed at two to investigate the environmental complexity, and they were located next to an outer wall and faced open space where a speaker array and movable/replaceable wall(s) were placed. The complexity of the environment was changed by varying two parameters of the movable/replaceable wall: the distance of the wall to the edge of speaker array L_d and the length of the wall L_w. The shorter

the distance and/or the larger the length, the more complex the environment due to the increased number of reflections of the sound signal.

Speaker locations are shown in Figure 5(a) as blue crosses. A microcontroller controlled speakers so that each speaker sequentially emitted white noise for a programmed period. A set of ILDs for a wall setting were thus collected automatically. Once the ILDs were collected, the ability of the proposed approach was evaluated by emitting sound from a speaker at some location within the area of the speaker array and identifying the location in the form of an observation likelihood. This location was different than that of the speakers of the speaker array to demonstrate the ability of the proposed technique to identify the target at an arbitrary position.

Figure 5(b) shows the developed experimental system and the dimensions and other parameters used in the experiments are listed in Table 1. Sound was sampled at 8,192 frequency bins within the audible range to capture its behavior accurately. 54 speakers were aligned to cover the open space. The distance and the length of the wall were varied to introduce both lightly NLOS and heavily NLOS environments. The case of two walls ($n_w = 2$) was tested in addition to the single wall case to increase environmental complexity. Only the distance of the wall closer to the acoustic sensor was varied.



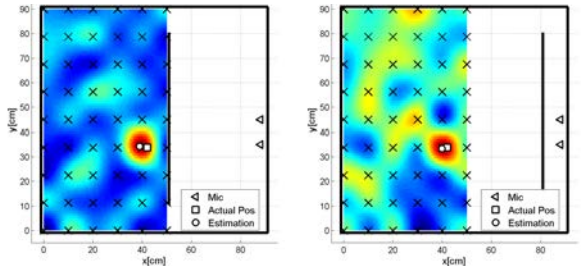
(a) Schematic design (b) Developed system

Fig. 5 Experimental system for investigating environmental complexity

Table 1 Dimensions and other parameters in the experiments

Parameter	Value	Parameter	Value
$\tilde{\mathbf{x}}^t$ single wall	$[42, 34]$ [cm, cm]	L	90 cm
$\tilde{\mathbf{x}}^t$ double wall	$[22, 56]$ [cm, cm]	Height	0 cm
ω_1	0 Hz	L_m	50 cm
ω_N	22 kHz	L_s	10 cm
N	8,192	L_d	$\{0, 10, 20, 30\}$ cm
ϵ	0.01	L_w	$\{50, 60, 70\}$ cm
n	54	n_w	$\{1, 2\}$

Figure 6 shows the resulting acoustic observation likelihoods when the sound target was at position $[42, 34]$ and $[22, 56]$ for the single wall and double wall cases, respectively. The former two cases were with a single wall at different distances. The latter two cases were with two walls with different wall length. The result first indicates that the target location is well estimated when the distance is short or when the length is small. The target is closer to LOS in these conditions since sound reaches the acoustic sensor with a small number of reflections. The identification of the target location in the remaining two cases is difficult due to the number of sound reflections. The identification with two walls is seen to be significantly harder than that with a single wall for the same reason. While the acoustic observation likelihood is heavily multi-modal with these cases, the target location is still captured by the highest peak or at least by one of the peaks as shown in Figure 6(d). This demonstrates the ability of the proposed approach to identify the location of the NFOV target though with limited accuracy.



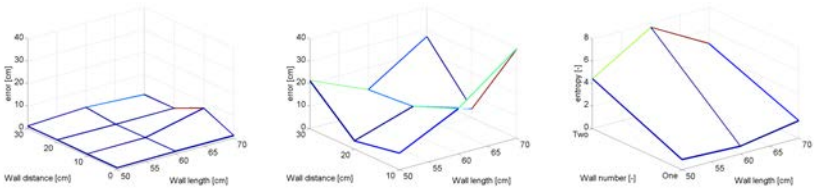
(a) $\{L_d, L_w, n_w\}$={0, 70, 1} (b) $\{L_d, L_w, n_w\}$={30, 70, 1}

(c) $\{L_d, L_w, n_w\}$={20, 50, 2} (d) $\{L_d, L_w, n_w\}$={20, 70, 2}

Fig. 6 Acoustic likelihoods for different environmental complexity

Figure 7(a) and 7(b) show the mean error of the acoustic observation likelihood when the distance and the length were varied for single and double wall cases. The mean error is the distance of the nearest peak of the acoustic observation likelihood to the true target location. The result of the mean error shows that the proposed technique could locate the target to within 2 cm error in 11 of the 12 cases for single wall case. The estimation was particularly good when the wall length was small. Figure 7(c) shows the uncertainty comparison for the two cases, using the differential entropy derived at a point within the normalized likelihood is used as the

uncertainty. The mean entropies for the two cases show that uncertainty increases with increase in number of walls for all wall lengths as expected. For the double wall case, the uncertainty is higher with less success in target identification, but the proposed approach could still be used to identify the target location.



(a) Single wall mean error (b) Double wall mean error (c) Differential entropy for single and double wall

Fig. 7 Mean error and differential entropy of the acoustic observation likelihood with a single and double wall

4.2 Applicability to Practical Indoor Scenario

4.2.1 Practical indoor scenario

Having validated the ability of the proposed acoustic sensing technique, the applicability of the proposed approach in NFOV target estimation to a practical indoor scenario was investigated. Figure 8 shows the actual indoor environment used for the investigation: the apartment of an elderly person who needs home health care service. As shown in the figure, the environment with five separate rooms is so complicated that it is difficult to cover the entire area by cameras. In addition, this is personal home, so cameras are not to be installed. The approximate dimensions of the apartment are 7.1 m in width, 10.4 m in length and 2.5 m in height. Six microphones, shown as red dots, were fixed to cover the entire space. The target person carried a small speaker which emitted sound with white noise. Parameters used for acoustic target estimation are listed in Table 2.

Table 2 Dimensions and other parameters in the experiments

Parameter	Value	Parameter	Value
ω_1	0 [Hz]	Height	5[cm]
ω_N	2.7 [kHz] k	n	255
N	2,000	ϵ	0.01
δ_S	2		

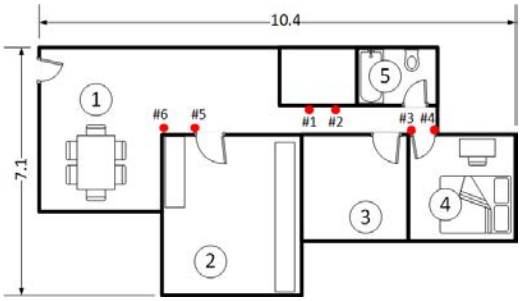
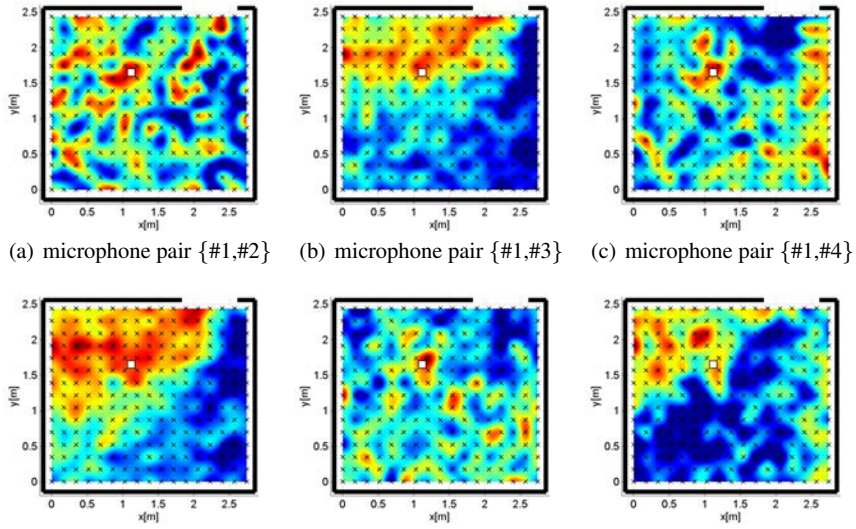


Fig. 8 Map of the test environment dimensions[m] and other details

4.2.2 Results

Figure 9 shows the acoustic observation likelihoods created by microphone pairs when the target person walked in Room 3. The square dot indicates the true target position. Only the likelihoods with microphones 1-4 are shown since those with microphones 5 and 6 did not meet the δ_{SNR}. Identified best of the combinations are pairs 2,3 and 1,3. Microphones 1-3 have the most direct LOS to Room 3, so the result matched well with the expected observable region. Figure 10 shows the resulting joint likelihood. The target location is accurately identified by filtering uncertainties.



(a) microphone pair {#1,#2} (b) microphone pair {#1,#3} (c) microphone pair {#1,#4}

(d) microphone pair {#2,#3} (e) microphone pair {#2,#4} (f) microphone pair {#3,#4}

Fig. 9 Acoustic likelihood in room 3 from multiple sensor combinations

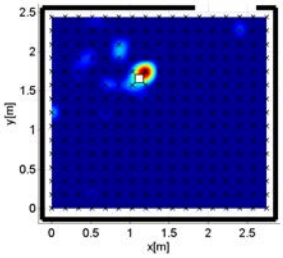
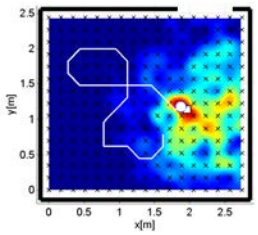
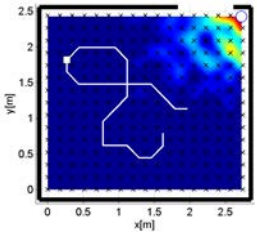


Fig. 10 Joint acoustic observation likelihood

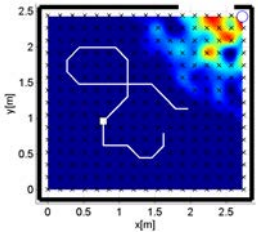
The result of RBE when the target person walked around is shown in Figure 12 with the true position again indicated by a square dot. It is seen that the proposed approach accurately tracks the target. The estimated position was less than 15 cm from the true target position in 83 % of the time. Cameras and RF receivers/transmitters cannot be used for such a highly constrained environment, so the conventional acoustic sensing technique based on two microphones was tested as the only comparable approach. As shown in Figure 11, the conventional approach was not able to identify the target location once it had failed in the localization.



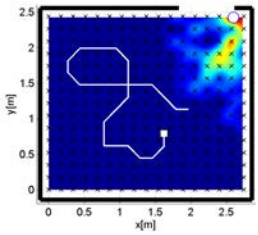
(a) $k = 1, [1.96, 1.13]$



(b) $k = 11, [0.27, 1.82]$

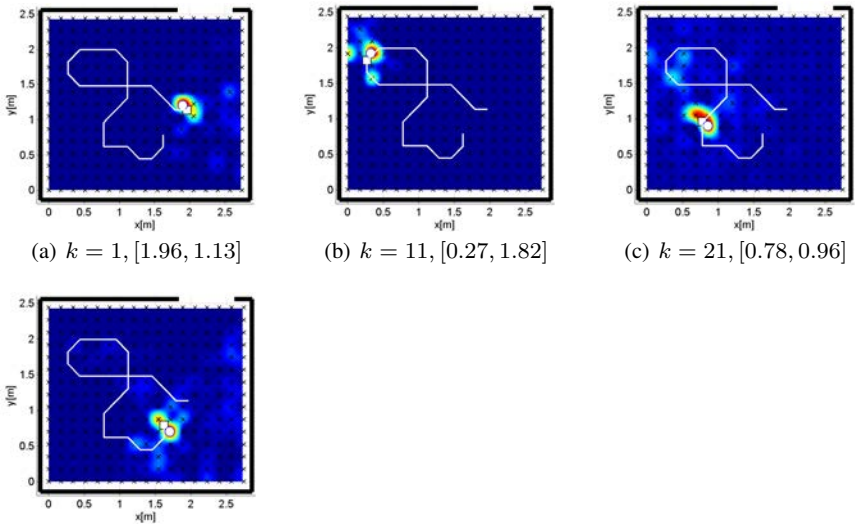


(c) $k = 21, [0.78, 0.96]$



(d) $k = 29, [1.63, 0.79]$

Fig. 11 Acoustic observation likelihood in room 3 with one microphone pair



(a) $k = 1, [1.96, 1.13]$ (b) $k = 11, [0.27, 1.82]$ (c) $k = 21, [0.78, 0.96]$

(d) $k = 29, [1.63, 0.79]$

Fig. 12 Proposed Joint acoustic observation likelihood in room 3 with RBE

5 Conclusions

This paper has presented a new approach which uses a set of microphones to localize and track a mobile NFOV target, and its applicability and implementation in complex indoor environments. The proposed approach derives the ILD of observations from a selected set of microphones for different target positions and stores the ILDs as acoustic cues. Given a new sound, an acoustic observation likelihood is computed for each pair of microphones by correlating ILDs. The joint likelihood is then created by fusing the acoustic observation likelihoods, and the NFOV mobile target is estimated by the RBE. Following the experimental parametric studies, the proposed approach was applied to track an elderly person needing home health care service, yielding an estimation which was successful to within 15 cm accuracy at 83 % of all the tested positions. These results have conclusively demonstrated the potential of the proposed approach for practical target localization.

The paper has demonstrated the new concept, and many challenges are still open for future study. The issues of immediate interest include the enhancement of acoustic sensing using the Interaural Time Difference (ITD) and the Interaural Phase Difference (IPD) as well as the use of non-white noise sound with sound separation/speech recognition techniques, so that the approach could be used for various applications. For the ILD database in a dynamic environment, automated update needs further investigation.

References

1. P. Bahl and V. N. Padmanabhan. Radar: An in-building rf-based user location and tracking system. In *INFOCOM 2000. Nineteenth Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings. IEEE*, volume 2, pages 775–784. Ieee, 2000.
2. J. Chen, J. Benesty, and Y. Huang. Time delay estimation in room acoustic environments: an overview. *EURASIP Journal on applied signal processing*, 2006:170–170, 2006.
3. Z. G. X. Dai, Huan Zhu. Multi-target indoor localization and tracking on video monitoring system in a wireless sensor network. *Journal of Network and Computer Applications*, 2012.
4. T. Furukawa, F. Bourgault, B. Lavis, and H. DurrantWhyte. Recursive bayesian search-and-tracking using coordinated uavs for lost targets. In *Robotics and Automation, 2006. ICRA 2006. Proceedings 2006 IEEE International Conference on*, pages 2521–2526. IEEE, 2006.
5. T. Furukawa, L. C. Mak, H. Durrant-Whyte, and R. Madhavan. Autonomous bayesian search and tracking, and its experimental validation. *Advanced Robotics*, 26(5-6):461–485, 2012.
6. I. Guvenc and C. Chong. A survey on toa based wireless localization and nlos mitigation techniques. *Communications Surveys & Tutorials, IEEE*, 11(3):107–124, 2009.
7. H. M. Khoury and V. R. Kamat. Evaluation of position tracking technologies for user localization in indoor construction environments. *Automation in Construction*, 18(4):444–457, 2009.
8. D. Kimoto and M. Kumon. Optimization of the ear canal position for sound localization using interaural level difference. 36th Meeting of Special Interest Group on AI Challenges, 2012.
9. M. Kumon, D. Kimoto, K. Takami, and T. Furukawa. Bayesian non-field-of-view target estimation incorporating an acoustic sensor. In *Intelligent Robots and Systems (IROS), 2013 IEEE/RSJ International Conference on*, pages 3425–3432. IEEE, 2013.
10. A. M. Ladd, K. E. Bekris, A. P. Rudys, D. S. Wallach, and L. E. Kavraki. On the feasibility of using wireless ethernet for indoor localization. *IEEE Transactions on Robotics and Automation*, 20(3):555–559, 2004.
11. L. Mak and T. Furukawa. Non-line-of-sight localization of a controlled sound source. In *Advanced Intelligent Mechatronics, 2009. AIM 2009. IEEE/ASME International Conference on*, pages 475–480, 2009.
12. R. Mauler. *Recent Developments in Cooperative Control and Optimizatio*, chapter Objective Functions for Bayesian Control-Theoretic Sensor Management, II: MHC-Like Approximation, pages 273–316. Kluwer Academic Publishers, Norwell, MA, 2003.
13. N. B. Priyantha, H. Balakrishnan, E. D. Demaine, and S. Teller. Mobile-assisted localization in wireless sensor networks. In *INFOCOM 2005. 24th Annual Joint Conference of the IEEE Computer and Communications Societies. Proceedings IEEE*, volume 1, pages 172–183. IEEE, 2005.
14. T. W. Sederberg, J. Zheng, A. Bakenov, and A. Nasri. T-splines and t-nurccs. In *ACM transactions on graphics (TOG)*, volume 22, pages 477–484. ACM, 2003.
15. C. K. Seow and S. Y. Tan. Non-line-of-sight localization in multipath environments. *Mobile Computing, IEEE Transactions on*, 7(5):647–660, 2008.
16. P. Svaizer, A. Brutti, and M. Omologo. Environment aware estimation of the orientation of acoustic sources using a line array. In *Signal Processing Conference (EUSIPCO), 2012 Proceedings of the 20th European*, pages 1024–1028. IEEE, 2012.
17. Y. Tamai, Y. Sasaki, S. Kagami, and H. Mizoguchi. Three ring microphone array for 3d sound localization and separation for mobile robot audition. In *Intelligent Robots and Systems, 2005.(IROS 2005). 2005 IEEE/RSJ International Conference on*, pages 4172–4177. IEEE, 2005.
18. D. B. Ward, E. A. Lehmann, and R. C. Williamson. Particle filtering algorithms for tracking an acoustic source in a reverberant environment. *Speech and Audio Processing, IEEE Transactions on*, 11(6):826–836, 2003.