This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

ReLiSCE: Utilizing Resource-Limited Sensors for Office Activity Context Extraction

Homin Park, Jongjun Park, Hyunhak Kim, Jongarm Jun, Sang Hyuk Son, *Fellow, IEEE*, Taejoon Park, *Member, IEEE*, and JeongGil Ko, *Member, IEEE*

***Abstract*—The capability to extract human activity context in a room environment can be used as meaningful feedback for various wireless indoor application systems. Being able to do so with easily installable resource-limited sensing components can even further increase the system's applicability for various purposes. This paper introduces our efforts to design a system consisting of heterogeneous low-cost, resource-limited, wireless sensing platforms for accurately extracting the human activity context from an indoor environment. Specifically, we introduce Resource Limited Sensor-based activity Context Extraction (ReLiSCE), a system consisting of microphone array, passive infra-red (PIR), and illumination sensors that effectively detect the activities that occur in an office (meeting room) environment. The signal processing schemes used in ReLiSCE are designed so that their size and complexity is suitable for the resource limitations that many embedded computing platforms introduce. Using empirical evaluations with a prototype system, we show that despite the simplicity of its data processing schemes, ReLiSCE successfully classifies human activity states in various meeting scenarios. Furthermore, we show that high accuracy is achieved by combining results from heterogeneous sensors. We foresee this paper as a sub-system that interconnects with various application systems for autonomously configuring people's everyday living environments in a more comfortable and energy-efficient manner.**

***Index Terms*—Computers and information processing software, context extraction, embedded software, low-power signal processing, smart environments.**

I. Introduction

With the vast development of wireless sensing and low-power networking technologies over the last decade, it is no surprise to us that these sensors will soon be deployed in our everyday living environments for various application purposes. A major subset of these systems are often applied to buildings as a way to increase the user comfort level, and efficiently control the growing energy usage expenses for building management. Applications in this domain include access control systems, smart lighting management systems, and heating, ventilation and air conditioning (HVAC) systems [1], [2]. These systems utilize different sensing platforms to monitor the target environment and report sensing results to an actuation unit, which controls the environmental parameters to satisfy the users' or applications' needs. Until now, the sensing components in these systems were simple, and raw sensor readings were enough to make an estimation of the current environment state. Nevertheless, in order to make these systems smarter, wireless sensing systems are now expected to make situation-aware decisions that are more human-centered than simply identifying low-level physical characteristics of an environment. Such smarter wireless sensing systems should be able to extract and determine the activity context in an indoor environment to adjust environmental configurations based on more complex reasoning.

However, when a system involves a level of computational complexity with the sensing data, typically until now, a server-scale computer that aggregates all the sensing information was in-charge of the decision-making process. For example, when using microphone array sensors to capture human activity using the collected acoustic signals, a PC-scale device was installed to execute the signal processing algorithms. As a result, to apply complex activity recognition-based systems, the installation of this additional hardware unit was essential. Such difficulties can potentially act as a barricade, which blocks the large-scale deployment of these smart wireless sensing systems. Nevertheless, if we can bring down such complex signal processing capabilities to operate on smaller-sized, resource-limited platforms, the applicability of these systems will greatly increase to cover various practical applications. While the prices of powerful processing units have come down over the last few years, for active commercialization of these systems, it is important to identify lowest-cost possible processing platforms, as long as we can configure the platform to provide services with the same quality (or with minimal application-level impact) as high-cost devices.

This paper targets the problem of designing an accurate activity recognition system for indoor office environments using resource-limited computing platforms and low-cost sensing components. Specifically, we utilize a microphone array sensing platform with optional supplementary sensing components, such as passive infra-red (PIR) and illumination sensors, to extract the current activity state of people in an office environment. Our system, Resource Limited Sensor-based activity Context Extraction (ReLiSCE), utilizes the aforementioned sensors and combines them with low-resource demanding

Manuscript received May 15, 2014; accepted August 15, 2014. This work was supported in part by the Daegu-Gyeongbuk Institute of Science and Technology, Research and Development Program of Ministry of Science, ICT and Future Planning (MSIP) of Korea (CPS Global Center), and in part by the IT Research and Development Program of MSIP/KEIT Project 10035570. This paper was recommended by Associate Editor Y. Xiao. *(Corresponding author: JeongGil Ko.)*

H. Park, S. H. Son, and T. Park are with the Daegu-Gyeongbuk Institute of Science and Technology, Daegu 711-873, Korea.

J. Park, H. Kim, J. Jun, and J. Ko are with the Electronics and Telecommunications Research Institute, Daejeon 305-700, Korea.

Color versions of one or more of the figures in this paper are available online at http://ieeexplore.ieee.org.

Digital Object Identifier 10.1109/TSMC.2014.2364560

2168-2216 © 2014 IEEE. Personal use is permitted, but republication/redistribution requires IEEE permission.
See http://www.ieee.org/publications_standards/publications/rights/index.html for more information.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

schemes for local, internal data processing. For the most resource-demanding sensor, the microphone array, we design a series of lightweight signal processing schemes, which include a signal peak detection scheme with time frame calibration, time difference of arrival (TDOA) computation, and angle-based speaker location estimation schemes. The results are combined with data from other sensors to make a near-perfect estimation on the human activity states.

We evaluate ReLiSCE using experimentally collected traces and also through a prototype deployment of the system in an actively used meeting room environment. Results from our evaluations show that, by combining the capabilities of heterogeneous sensors, ReLiSCE successfully detects the activity context of the room environment with high reliability.

Specifically, we summarize the main contributions of this paper in threefold.

1) We propose ReLiSCE, which utilizes an embedded microphone array sensor running novel lightweight signal processing schemes, along with the assistance of heterogeneous sensing modalities to accurately and reliably estimate current human activity contexts in a room environment.
2) We discuss the technical challenges that complicate the design and integration of various sensor processing algorithms when utilizing resource-limited, low-power embedded platforms. By doing so, we introduce various novel schemes, and especially detail our discussions for the microphone array sensor, which in previous work has mostly been integrated with high-end (e.g., PC-scale) hardware.
3) Using a large set of data samples collected from real meetings and a prototype system implementation deployed in real meeting scenarios, we show that the performance of ReLiSCE is satisfactory enough to be quickly applied to various applications and showcase its effectiveness using a smart lighting application.

The remainder of this paper is structured as follows. In Section II, we introduce the importance of office-level activity context extraction for various application scenarios. Next, in Section III, we introduce a set of activity context scenarios that this paper targets and also detail the signal processing schemes in ReLiSCE, a distributed data processing system that effectively combines readings from multiple heterogeneous sensors to make a decision on the activity context of a target environment. We evaluate the performance of ReLiSCE in Section IV using a month-long deployment of our system in a real meeting room environment. Section V positions this paper among other related literature and we summarize this paper in Section VI.

II. Activity Context Extraction: Issues With Utilizing Resource Limited Hardware

As wireless sensing systems start to permeate closer to our everyday living environments and automate many routine tasks that were previously manual [3], system designers now target to seek a higher goal of automatically detecting people's activity context and reacting accordingly to configure an even more comfortable environment. For example, by knowing what activity is occurring in an office/meeting room environment, the lighting conditions can change autonomously without manual user controls. In designing such systems, a major requirement (and challenge) to address is the accurate detection and estimation of human intentions in the target sensing field. To achieve this goal, many, as we will later discuss in Section V, have proposed systems for extracting human activity context using various sensing components. However, since such systems deal with high data rate sensing and processor intensive data processing, PC-scale devices are typically used in the data processing phase. While this is a reasonable system design, the capability to perform such operations internally at the sensing unit can effectively widen the application domain where such activity context extraction systems can benefit. Once we can embedded this computing capability to smaller-sized computing platforms, utilizing these systems will no longer require the installation of a designated PC, and will be easier to commercialize since embedded computing and sensing components can be sold as commercial off the shelf (COTS) devices. Still, the problem is that many activity context extraction algorithms require a level of computational-intensive operations using high data rate sensor samples. Therefore, allowing these algorithms to execute on resource-limited embedded platforms, is still a challenge. One example of such complex algorithms is the signal processing schemes used for extracting human activity with microphone sensors. While vocal acoustic signals hold comprehensive information, which can be extracted to estimate the current human activity context, powerful processors are typically needed to collect and process these signals for extracting meaningful information.

With schemes that are lightweight and effective enough to accurately detect the human activity context, system designers can potentially propose application systems that are a step more user-friendly and ambiently intelligent. Furthermore, going back to the microphone example, being able to locally process the data on embedded platforms diminishes the need to transmit raw sensor measurements (e.g., acoustic signals sampled at high frequencies) to PC-scale devices; thus, allows the sensors to be easily deployed wirelessly.

We point out that over the recent years, the processing power of embedded platforms have increased dramatically. Nowadays, the smartphones that we carry around embed quad-core processing units, which is (more than) enough to perform computation-intensive operations. However, applying such high-end processors to nongeneral commercial products will increase the product cost and reduce the system's commercial attractiveness. Therefore, we argue that a cheapest-possible resource-limited hardware platform with fully optimized software components should be favored for extracting human activity context. While ideal, we detail some challenges and milestones required in designing such systems below.

1) *Identifying the Target Activities:* Based on the goals of the application system, system designers should accurately identify the types of activities they intend to detect. The activity states generated from humans can be diverse and it is easy to say that not all of these states can be discretely defined. As a result, it is important to set a goal on what activity states the system is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

interested in extracting. For example, an indoor lighting control application system would benefit from knowing the conversational state in the target room environment while not requiring the knowledge of the meeting participants' current emotion status. Therefore, for each application, precisely defining the types of human activity states the system needs to detect is important.

2) *Integrating Heterogeneous Sensors:* The detection of human activity is complex, and is difficult to precisely predict using a single sensor module. Our experiences, as detailed in the following sections, show that data from heterogeneous sensing components should be integrated to make accurate predictions on the human activity states.
3) *Designing Less Resource Demanding Algorithms:* A decade of wireless sensing system research has contributed in proposing a number of low-resource demanding algorithms. Nevertheless, the applications where these systems were applied until now, mostly deal with sensors that are designed to be low-power. However, extracting human activity states can benefit from using more complex sensing components, in which no low-resource demanding algorithm exists. As a result, customizing the data processing algorithms so that they are capable of extracting the anticipated information is essential.

In this paper, we focus on extracting human activity states in an office and meeting room environment. Specifically, we design a base system that targets to provide user activity context for larger application systems such as smart lighting control or HVAC. The following sections provide details on how ReLiSCE, our proposed system for activity context extraction, addresses the aforementioned challenges and performs accurate human activity context extraction.

III. ReLiSCE: Utilizing Resource-Limited Sensors for Activity Context Extraction

When designing ReLiSCE, we carefully consider the aforementioned challenges. Nevertheless, the utmost goal in the system design is to enable embedded, resource-limited computing platforms, instead of PC-scale devices, to serve as the data processing and decision making component of our system. We especially put most of the focus on the signal processing component for the microphone array sensor, which possesses the most challenges. While other sensors in ReLiSCE can be applied to resource-limited platforms using simple algorithm-based schemes (as detailed in the subsections that follow), each microphone module in a microphone array outputs a series of digitized acoustic signals, which require further (extensive) processing to reform the raw data into useful information. In this section, we detail the hardware components in ReLiSCE, and present reasons for selecting such modules along with a set of (low resource-demanding) schemes that achieve the application-level goals of accurately extracting human activity states.

A. Defining the Target Detection Activity States

We first start our discussions by introducing the target activity contexts that we expect ReLiSCE to accurately detect. This paper of activity context extraction for meeting environments, initiated in the 1980s, has been widely studied for decades. The objective of these studies were to understand and classify different group actions using extracted behavioral information. In McGrath's [4] task circumplex model, group actions were divided into eight distinct categories depending on how individuals interact in a group. This categorization model includes the perspectives of social psychology.

We note that for computational realm, the definition of each action must be much more constrained than that of McGrath's while its insights are well preserved. Considering this, McCowan *et al*. [5] defined a set of high-level group actions including: monolog, discussion, presentation, white-board, and note-taking. We detail each action state as follows.

1) Monolog represents the activity status where an individual speaks continuously without any interruption (from other meeting participants) for a certain period of time. Such an action can be observed when someone tries to explain his or her idea to the group.
2) In contrast, the discussion state happens when multiple individuals participate to speak up during a certain time period in order to share ideas.
3) Presentation is defined as a state when the projector is utilized during a meeting while an individual (i.e., a main presenter) speaks up.
4) When the white-board is used rather than the projector, the group is said to be in the white-board action state.
5) When individuals in a group start to write a note with no verbal activity, the system defines such action as note-taking.

With the objective of this paper, where resource-limited sensing platforms are favored than high-end devices, we are highly interested in the set of actions defined above, excluding the note-taking and white-board states. While verbal and device utilization cues can be acquired using comparatively simple devices and algorithms, detecting hand gestures for these two states would require computationally heavy video processing and pattern recognition processes.

We further extend our detection states with a room-idle state, a room-enter state, and a room-in-use state, indicating that the meeting room is no longer used, that the users initially enter the target environment (instantaneous state) and that the room is continuously in-use, respectively.

B. Hardware Components in ReLiSCE

We now provide details on the hardware components in ReLiSCE and provide the reasons behind selecting such components with respect to the target detection activity states mentioned above.

1) Microphone Array Sensor: At the core of ReLiSCE is the microphone array sensor as pictured in Fig. 1. On the application's perspective, the microphone array sensor can be used to effectively help distinguish the monolog and discussion states by extracting information on how many speakers are active in the environment. Our microphone array sensor consists of four low-power consuming microphone modules with amplification capabilities (up to 40 dB), each connected to an ADC port of a Mango Z1 Board [6]. The Mango board is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

Fig. 1. Picture of our prototype hardware implementation of the microphone array sensor. The processing unit is an ARM cortex M3-based STM32 microcontroller and four microphone sensors with amplifiers are connected via the ADC ports.

a computing platform that combines an ARM cortex M3-based microcontroller (MCU) with an IEEE 802.15.4 2.4 GHz radio. We note that the diagonal length of the microphone array sensor prototype is ∼45 cm.

While the schemes that we will discuss in Section III-C is capable of operating under even-restricted MCUs, we use a (relatively) powerful ARM cortex M3-based MCU due to the fact that MCUs with more resource restrictions (e.g., Atmel 1284p and TI MSP430) cannot support the high sampling rates of the microphone array. Based on the fact that these microphones target to detect human voice-originated acoustic signals, we target to sample each microphone at 44 kHz with sample sizes of ten bits. This would mean that, given four microphone sensors in an array, the sampling rate at the MCU must be >176 kHz. Our MCU allows the latency of collecting a single sample to fall <0.07 μs, which is much shorter than the latency bound of capturing the same vocal (acoustic) characteristics at the four microphones. Unfortunately, widely used MCUs such as the Atmel ATmega 1284p can only support ADC sampling up to 15 kHz (e.g., 200 kHz ADC clock timer with 13.5 cycles per conversion); thus, is not sufficient enough to support the microphone's sampling requirements.

In ReLiSCE, the processing module at the microphone array sensor also takes the role of decision making for human activity extraction. It aggregates input from different (wirelessly connected) sensing components to derive the estimated human activity state.

2) PIR Sensor: While the microphone array distinguishes different types of conversation, the acoustic signal-based nature of the microphone limits its capabilities for detecting other activity states without vocal communication. Examples of such are the states related to room occupancy (e.g., room-idle, room-enter, room-in-use). Based on a survey of different sensing modalities, we noticed that the PIR sensor was the ideal candidate for detecting such activities. For this reason, ReLiSCE includes a PIR sensor with a detection angle of 120 degrees and a detection distance of ∼6 m. The sensor is attached to an Atmel ATmega 1284p MCU and a TI CC2420 IEEE 802.15.4 radio. In operation, the PIR sensor detects human existence in a target region and periodically (or on request) reports its results to the microphone array sensor's computing module for activity state estimation.

3) Illumination (Light) Sensor: When the microphone array determines that the environment is in the monolog state, ReLiSCE makes sure if the single speaker is simply in a conversation or if the room is in a presentation state. However, it is difficult for ReLiSCE to distinguish between the two states simply using the microphone array. We note that with predefined locations of the presentation stage, we can potentially use the speaker's angular locations to estimate the current activity context. While theoretically possible, in reality, it is difficult to assume that the microphone array will continuously be at the same location, since activities such as room cleaning can change its placement on the table. For this purpose, ReLiSCE also includes an illumination (light) level sensor with the goal of detecting beam projector activities. The illumination sensor is connected to either the projection screen (present in most meeting rooms) or near the light-bulb of the projector for accurate detection of the projector's activity. We emphasize that using the illumination sensor is a design choice that is targeted specifically for the detection of the presentation state. Nevertheless, we determined that applications such as smart lighting control can benefit from the accurate detection of this activity; thus, decided to add the illumination sensor as part of ReLiSCE. While we initially attempted to detect the projector's activity by identifying the sounds of its fan in operation, due to the different installation points of the projector (some were placed on the table and some were installed on the ceiling) the detection accuracy was very low. Similar to the PIR sensing platform, the illumination sensor is connected to an Atmel ATmega 1284p MCU and a TI CC2420 IEEE 802.15.4 radio to report data to the sensor data aggregation unit (e.g., microphone array sensor's computing component).

4) Eliminating the Camera Sensor: Using a camera sensor is an accurate way to detect the activity state of people in a target environment. Using image processing techniques, most of the operations in ReLiSCE would be easily achievable using existing image processing algorithms. However, despite its accuracy, camera sensors are mostly used in applications that relate to surveillance and safety due to the fact that in typical scenarios, they hold the potential of invading people's privacy. Furthermore, since most image processing algorithms require a level of computational power, applying them directly to the applications that we target in this paper is not a viable option. Therefore, this paper excludes the option of using a camera as a way to propose a privacy-sensitive, widely applicable human activity extraction system that operates on resource-limited embedded platforms.

C. Microphone Array Sensor Processing Scheme

For supporting the microphone array sensor in ReLiSCE, we analyze acoustic input signals from the microphones to detect the number of speakers within a moving window-based time interval t_α. To do so, we first perform a study on the processing requirements of existing microphone signal processing algorithms. In most existing works, the target is to compute the TDOA between each microphone pair by using the received acoustic signal patterns. These TDOAs are computed by identifying the time lag between each signal pair. In general, for environments with room reverberations and background noises, the generalized cross correlation with

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

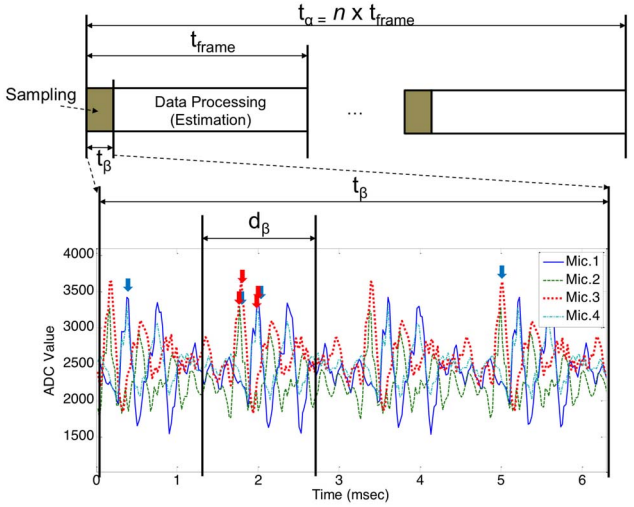


Fig. 2. Activity extraction process for the microphone array sensor. The processing, t_{frame} is divided in a sampling process and a data processing phase, t_β. d_β defines the short interval of acoustic signals considered from the data processing phase remaining after time frame calibration.

phase transform (GCC-PHAT) is widely used for TDOA computation [7].

In GCC-PHAT, a TDOA between each signal pair is computed by identifying the time lag Δd_{ij} that yields the highest correlation between a microphone pair ij. Using Δd_{ij}, a system locates the direction of the speakers, and uses this information to detect the number of speakers in the target environment. Such an algorithm would allow the system to distinguish between the monolog and discussion activity states. Specifically, with acoustic signals incoming from a single direction (only), the system can determine that the current activity state is "monolog," otherwise, in a "discussion" state when more than one speaker directions are detected within t_α. As the top of Fig. 2 shows, t_α can consist of multiple speaker estimation periods of which each period consists of duration t_{frame}. Nevertheless, while effective, the GCC-PHAT algorithm is a computation-hungry process that intensively uses fast Fourier transform (FFT) and cross-correlation computation. Such high resource-demanding characteristics of GCC-PHAT makes it inappropriate for use on resource-constrained embedded platforms.

In order to design a scheme that achieves the same results, but is compact enough to operate on resource-limited platforms, we aim to compute TDOAs by simply detecting the peaks of incoming acoustic signals from each microphone in the array for a (very) short time interval $t_\beta < t_{\text{frame}}$ (rather than matching the entire signal patterns as in GCC-PHAT). While detecting only the peaks reduces the computational complexity, two major challenges arise. First, it is unclear if simple peak detection is sufficient enough to assure the proper detection of speakers' directions. Second, despite the scheme being resource-efficient, we need to assure that it has enough time to execute while sampling the ADC (for the microphone array) at high rates. We first address the second question by proposing a microphone ADC duty-cycling scheme, in which the platform samples the ADC units for only a subset of t_{frame} (e.g., $t_\beta < t_{\text{frame}}$) and dedicates time for the signal processing scheme to execute. This not only allows the platform to exclusively allocate time for signal processing, but also reduces the memory usage burden that continuous high-frequency sampling can introduce to resource-limited platforms. Specifically, as the top of Fig. 2 shows, we divide the microphone-based detection process into two phases where the first phase focuses on vocal signal collection, and the second phase focuses on data processing. For any target microphone sampling rate λ [Hz], we note that with a platform consisting of k microphones, the main MCUs ADC sampling rate should support up to $k \times \lambda$ [Hz].

Given this two-phase process, our scheme, which is executed in the estimation phase of Fig. 2, operates in two steps. The former collects sound samples at each microphone for t_β and removes noise by passing signals through a low-pass filter, while the latter effectively and reliably estimates the direction of arrival (DOA) for the sound source by executing the three procedures as follows.

1) *TDOA Estimation Phase:* To efficiently estimate the TDOA, we first focus on computing the time of arrival (TOA) for the signals at each microphone. Instead of using resource-demanding FFT and cross-correlation, we introduce a signal peak detection scheme. Using the collected acoustic signals within time duration t_β, we compute the highest points of amplitude for the acoustic signals from each microphone (e.g., one highest peak per microphone; see the four red arrows on the bottom of Fig. 2). The intuition behind this is that a peak value of the acoustic signal initiated by the sound source at time instance t will be received at each microphone i at time d_i, which is computed as
$$d_i = t + t_i^{\text{propagation}}. \quad (1)$$
Here, $t_i^{\text{propagation}}$ is the propagation delay of the acoustic signal traveling at the speed of sound between the sound source and microphone i. Assuming that such peaks will be captured at all microphones of the array, we can guarantee the accuracy of TOA for each microphone by using the computationally-efficient peak detector as long as the effects of signal fluctuations (due to reflection, refraction, or noise) are minimal. Then, for each microphone pair (i, j) we can compute $\Delta d_{ij} = |d_i - d_j|$, which is the TDOA for (i, j). While technically a single MCU cannot simultaneously sample each microphone at the same time, since the MCU on our platform supports an ADC clock of up to 14 MHz and a conversion time of 12 clock cycles (e.g., 1.17 MHz ADC sampling rate), sampling all of the microphones on the array is faster than the time limit requirement of 44 KHz acoustic signals (for human voice). Thereby, all of the microphones (given that our microphone array sensor consists of four units) will sample the same (or very similar) time instances of the incoming acoustic signal.
2) *Time Frame Calibration Phase:* The aforementioned peak detection scheme works perfectly in an ideal environment. However, due to real environmental characteristics such as room reverberations and background noise, peaks may occur at spurious time points (i.e., deviating

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

significantly from true d_i's; see the four blue arrows on the bottom of Fig. 2). To address this practical issue, ReLiSCE includes a time frame calibration phase which operates as follows. Let $t_{\overline{ij}}$ denote the time for an acoustic signal to travel the distance between microphones i and j. Then, given a peak detected at microphone i at time t, the same peak will be captured by another microphone j within a time interval of $[t-t_{\overline{ij}} : t+t_{\overline{ij}}]$ since j can only be either closer to or farther away from the source. Given a square-shaped microphone placement with side length l (see Fig. 1), two microphones on the diagonal of the square yields the largest $t_{\overline{ij}}$ $(= \sqrt{2}l/c)$ where c is the speed of sound. Accordingly, the maximum TDOA for a given system can be derived as

$$\Delta d_{\max} = \frac{\sqrt{2}l}{c}. \tag{2}$$

Accordingly, any $\Delta d_{ij} > \Delta d_{\max}$ indicates that the TOA computation is invalid, or in other words, the two signal traces are observing different peaks. Based on this observation, we set $d_\beta = 2 \times \Delta d_{\max}$ and perform validations to confirm that the received TDOA samples are usable. As a result of the process, the original peaks detected at the blue arrows in the signal pattern of Fig. 2 are corrected as the red arrows within time frame d_β.

3) *Angle of Arrival (AOA) Computation Phase:* Finally, with a set of valid d_i's and Δd_{ij}'s, we compute the AOA for the sound origin. This is essentially the process of counting the number of speakers in the detection region. For doing so, we identify: 1) the microphone MIC_{first} that first detects the peak (and hence is closest to the sound origin); 2) another microphone MIC_{last} that detected the peak at last (farthest from the speaker); and 3) the microphone pair (i, j) with the smallest TDOA

$$\Delta d_{ij}^{\min} = \min\left(\forall_i \forall_j \Delta d_{ij}\right). \tag{3}$$

Furthermore, we denote a plane formed by a microphone pair i and j with $\Delta d_{ij}^{\min}$ as p_{ij}. Using $\Delta d_{ij}^{\min}$, we compute the AOA (in radians) with respect to p_{ij} as below

$$\phi_{ij} = \arccos\left(\frac{c \cdot w}{\sqrt{\left(x_i - x_j\right)^2 + \left(y_i - y_j\right)^2} \cdot \lambda}\right). \tag{4}$$

Here, w is the latency in the sample data, λ is the sampling rate, and x_i, y_i, x_j, y_j represent the coordinates of microphones i and j, respectively.

Using ϕ_{ij}, θ_{ij}, the DOA for the sound source, is computed in two different ways. First, we consider the case where MIC_{first} and MIC_{last} are located on different sides of plane p_{ij} (see left of Fig. 3). For such cases, $\theta_{ij} = \pi - \phi_{ij}$. For all other cases (see right of Fig. 3), $\theta_{ij} = \phi_{ij}$.

Using these steps, for each t_{frame}, we detect the plane p_{ij} and DOA θ_{ij} of the sound sources. Assuming that the speaker will not make significant movement while speaking, these parameters will remain at a single location during the speaking duration. Even if they are mobile, since t_β is short (e.g., few hundred ms), the detection of the speaker's location for the time instance should not be an issue. Nevertheless, with p_{ij} and θ_{ij} computed every t_{frame}, we check the validity of the result by checking if the same result is measured consecutively for k times within a θ_{ij} error-bound of $\pm\epsilon$.

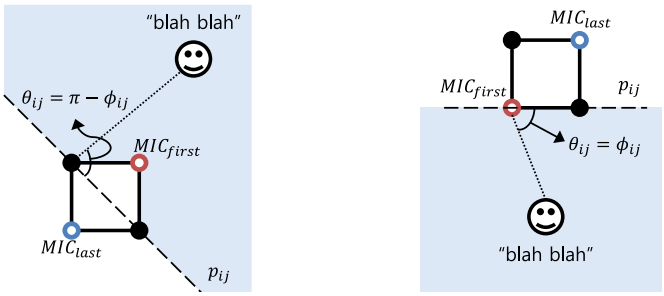


Fig. 3. Two methods of computing the DOA with respect to the relative location of the speaker to the microphone array sensor.

Lastly, based on these results, ReLiSCE enters the activity state classification stage. In this final process, ReLiSCE simply counts the number of results that are using different planes and are distantly positioned by more than ϵ degrees from each other. If this count is 1, then this is an indicator that a single person is actively speaking in the room, which allows the system to conclude the room is in a monolog state. In the case where more than one speaker positions are detected, ReLiSCE sets the activity state of the room to discussion. We note that there is a possibility that if multiple speakers are located within $\pm\epsilon^\circ$, ReLiSCE can misclassify the room activity as monolog. Fortunately, in the meetings traces that we collected, such an error was not observed. Nevertheless, this is a potential limitation that ReLiSCE possesses and we plan to address this in future work by analyzing the vocal patterns of different speakers to more accurately separate speakers.

D. PIR Sensor Processing Scheme

Compared to the signal processing operations for the microphone array sensor, the data processing for other sensors in ReLiSCE are relatively simple. Our PIR sensor outputs a low-to-high edge on the connected GPIO when an object is present in the target detection region. A series of the collected output values from the PIR sensor is passed through a filter to minimize the detection of false positives. This comes from empirical observations of multiple high values on the PIR sensor rather than a single high-edge observation when a person is present in the detection field. Using such series of "PIR detected" patterns reported from the PIR sensor, ReLiSCE defines the room-enter state and continues itself in the room-in-use state.

While the PIR sensor is effective for detecting the room-enter state, due to its sensitivity, detecting idle activity can be a convoluted process. For example, the PIR sensing platform producing an output of "no detection" can both imply an empty/idle state or an "active" state where people are still in the room with no active movements. Therefore, it is important to set a time interval before the PIR sensing platform determines that the room is in the room-idle state. ReLiSCE

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

introduces an adaptive timer for this purpose based on the following.

1) Once the room is in room-enter state, the PIR sensor keeps continuous track of the activity levels in the room with respect to its PIR readings.
2) The idle time determination duration T_{idle} is set to an initial value and frequent PIR high detections naturally decrease this timer value linearly. Here, $T_{\text{idle}} > T_{\text{idle}}^{\min}$, where $T_{\text{idle}}^{\min}$ is a predefined as the minimum wait duration for idle state determination.
3) If no PIR high measurements are detected for T_{idle}, the PIR sensor determines that the system in a room-idle state. This design decision comes from empirical experiences showing that a meeting with large amounts of active movements tend to continuously have such characteristics, as this is an effect of the people participating in the meeting.
4) The PIR sensor data is analyzed discretely on a per-meeting basis (e.g., the time is reinitiated when ReLiSCE determines a new room-enter state) and ReLiSCE infers a proper minimal timing value for each meeting to determine the room-idle state.

E. Illumination Sensor Processing Scheme

The scheme for processing the illumination sensor focuses on exploiting an adaptive feature that our sensor, and many modern light level sensors, introduce. Specifically, we exploit the fact that these sensors provide different modes of detecting the illumination levels with respect to the ambient brightness of the surrounding environment. In other words, by specifying whether the sensor should be in the high- or low-gain mode, the same number of sampling bits (e.g., 10 bits on the ADC) are mapped into 0–500 lux in high gain mode, and 0–50 000 lux in low-gain mode [8]. These ranges can change with respect to the electronical details, but, in any case, a high-gain mode, allows the illumination sensor to be more reactive and sensitive in darker environments while the low-gain mode enables detection over wide light intensities.

In ReLiSCE, we start sensing the illumination sensor in the high-gain mode, to detect the initial activities of the projector unit. When the projector is on, or the room is fully lighted, we adaptively change the hardware configurations to low-gain mode so that our sensor can make accurate measurements of the lighting conditions. Based on the collected illumination data, we utilize a Δ threshold-based scheme, where we compare the exponentially weighted moving average of the previous samples to the current sample to capture changes in the lighting levels (e.g., activity of the beam projector).

F. Combining Heterogeneous Sensors in ReLiSCE

While distinguishing monolog and discussion states is simple when using the decision made from the microphone array sensor, ReLiSCE combines the capabilities of the other two sensors (with the results from the microphone array) to make comprehensive decisions on other activity states. This section, and the state diagram illustration in Fig. 4, provides details

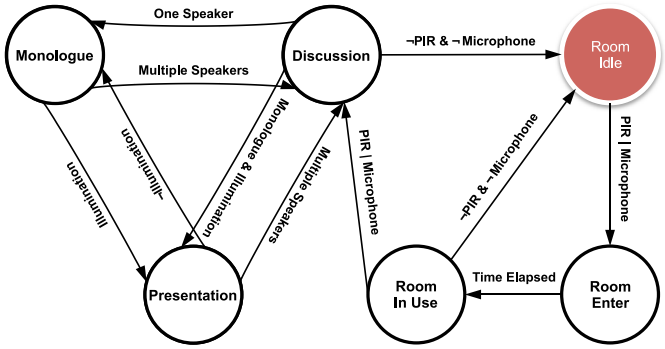


Fig. 4. State diagram of determining activity states with heterogeneous sensors in ReLiSCE.

on how heterogeneous sensing modalities in ReLiSCE are combined to output a single system-level decision.

1) Detecting Room-Enter and Room-Idle States: For detecting the room entering state, we combine the capabilities of the microphone array and the PIR sensing platform. We note once again that all the decisions are made at the computing platform on the microphone array sensor, since it is equipped with (relatively) the most computing resources (e.g., a higher power MCU). Specifically, once this decision-making component collects the current PIR state (periodic or on-demand basis), it determines if the PIR sensor reports an "in-use" state. At the same time, it also checks if the microphone array is detecting a monolog or discussion state. If either (e.g., `OR` function) of these states show that the room is being used, then ReLiSCE declares a room-enter state and notifies the application system (e.g., a smart lighting system or HVAC) that someone had entered the room. Shortly after this, ReLiSCE declares that the room is in the room-in-use state, and initiates the meeting room context extraction schemes. The time between room enter and room-in-use states is determined at the time of deployment. We currently set this to 1 s. The main reason for this separation is to prepare time for the other sensors to initiate their settings in case these sensors are in a low-power mode.

As for detecting the room's idle state, while we utilize the same set of sensors, ReLiSCE declares the room to be in the room-idle state if both (e.g., `AND` function) sensing components report inactivity. We make such a design choice given that most smart environment applications are required to react more sensitively when the room is in use to ensure a satisfactory user-experience; therefore, requires a more conservative behavior when declaring room-idle.

2) Detecting Presentation State: For the presentation state, ReLiSCE expects the system to first be in an active state (e.g., all states but room-idle). Note that, we focus on detecting presentation by combining the knowledge of the current speaking state (using the microphone array) and the projector's power state (using the illumination sensor). Specifically, with the results from the microphone array, the decision making unit first checks if the room is in the monolog state. This is an indication that someone is dominant in the speaking activities within a meeting. Furthermore, if the illumination sensor indicates that the projector is currently active, ReLiSCE declares the presentation state and notifies the application system of

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

this decision. Accordingly, applications such as smart lighting systems can automatically adjust its illumination levels to optimize for the viewing of the projection screen.

G. Minimizing Sensor Usage

While, we envision that the sensors in ReLiSCE, especially the microphone array, will operate with its power supplied from a wall-plug, we still expect the platforms to minimize their computing resources utilization. The fact that a subset of ReLiSCE's sensing components can (potentially) operate on battery (e.g., for easy installation), makes such considerations important. For example, it is easier to install a battery-powered illumination sensing component to a projection unit than to find additional power-plugs for this single sensor.

ReLiSCE's resource conserving mechanism starts with detecting the room-idle state. Once entering the room-idle state, the illumination sensor is put to sleep and the radios on all sensing platforms enter a low-power mode. In this mode, the components' radios are turned off to minimize the idle listening times and they start operating a low-power medium access control layer, which is designed to be similar to low-power listening [9]. However, unlike the illumination sensor, the microphone and PIR sensors should not fully turn off their sensing capabilities, since they need to continuously monitor the environment for entering activities. Nevertheless, their radios are duty-cycled as well and the sensors are configured to sample the environment at lower rates. Moreover, for the microphone array, only a single microphone module is kept on to conserve power at the others.

IV. Evaluations

Based on the system description of ReLiSCE, we now present performance results from experimental evaluations on the effectiveness of each sensing component under different human activities.

A. Microphone Array Sensor Performance

To effectively showcase the performance of our microphone array's signal processing scheme, we implement the proposed scheme in two evaluation environments. Specifically, in addition to implementing the scheme on our hardware prototype, we performed a series of validation tests with a MATLAB-based implementation using real vocal acoustic signal data collected from our prototype hardware. We also use this MATLAB environment to make performance comparisons with the GCC-PHAT-based scheme used in several previous work. We do this comparison in MATLAB given that GCC-PHAT-based signal processing schemes are resource demanding; thus, cannot be implemented on our resource-limited hardware prototype.

We start our evaluation by comparing the detection performance of our proposed peak detection-based speaker locating system (with and without time frame calibration) against the performance of a GCC-PHAT-based algorithm in MATLAB. Using Figs. 5 and 6, we present the average angular error in detecting the speakers' DoA with varying sampling rates

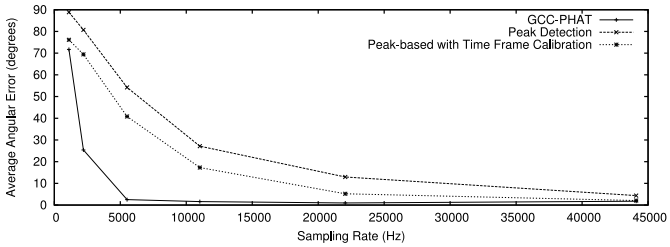


Fig. 5. Microphone array processing scheme accuracy with varying sampling rates.

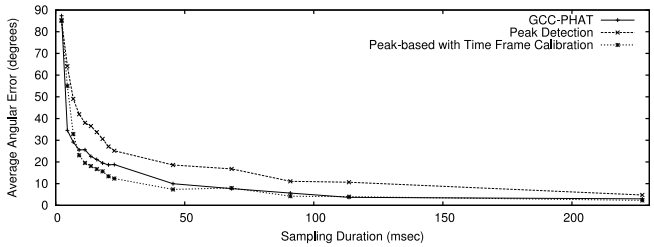


Fig. 6. Microphone array processing scheme accuracy with varying evaluation set sample durations.

and evaluation set sizes. Here, a low angular error would suggest that the number of speakers in the environment can be accurately estimated. The dataset used in this experiment was empirically collected from a single speaker at various relative locations of the microphone array sensor. For ground truth data, we keep track of the speaker's locations during data collection. Note that while we sample each microphone at 44 kHz for the data set, in Fig. 5, we intentionally down-sampled the data to observe the scheme's performance with lower ADC/microphone sensor sampling rates. We do this to validate the possibilities of using MCUs with lower sampling rates. Results in Fig. 5 suggest a few interesting points. Notice that the performance of GCC-PHAT outperforms our proposed peak detection-based scheme both with and without time frame calibration. This is even more pronounced when the sampling rate is low, which is an expected result given that the cross-correlation functions can be effective in these cases. For the two peak detection-based schemes, the sampling rates need to reach a higher level since each microphone should be able to detect the same peak for the incoming acoustic signals. As a result, as the sampling rate reaches 44 kHz, the detection accuracy of our proposed scheme matches that of the GCC-PHAT algorithm.

By observing the performance with varying sample collection durations (e.g., t_β) with 44 kHz sampling in Fig. 6, we can notice once again that when $t_\beta >\sim 110$ ms, the average angular error of our proposed scheme (with time frame calibration) is $< 3\%$. This suggests that the microphone sensor can be operated with minimal performance overhead, while maintaining a low DoA error. Naturally, this indicates that the number of active speakers can be accurately detected; therefore, monolog and discussion states can be extracted in ReLiSCE with high accuracy.

As the final part of our MATLAB evaluations, in Fig. 7, we plot the complexity of the two methods (proposed versus GCC-PHAT) using the operation count observed in MATLAB.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

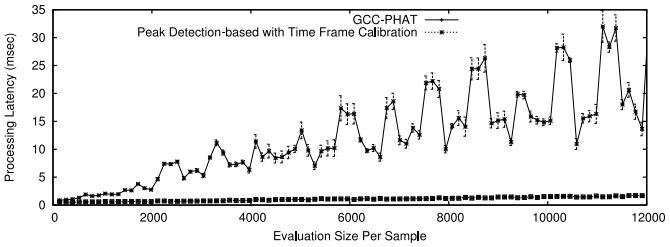


Fig. 7. Computational complexity (e.g., computation latency) comparison with ReLiSCE and GCC-PHAT.

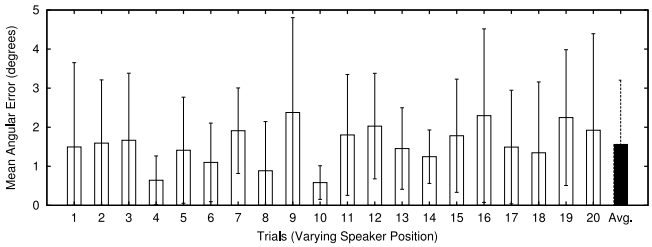


Fig. 8. Empirical speaker detection accuracy collected with ReLiSCEs microphone array processing scheme.

Notice that the increase in computing latency is gradual for the peak detection-based DoA estimation scheme, while GCC-PHAT shows a steep increase. This result is a clear evidence that ReLiSCE significantly lowers the computational burden at the processor, which is an essential factor when applying acoustic signal processing algorithms on resource-limited platforms. We point out that the fluctuation in the data for GCC-PHAT is mainly caused from its FFT component, which solves an n-point discrete Fourier transform (DFT) using divide-and-conquer. By recursively decomposing n-point DFT into smaller pieces, FFT reduces the number of complex multiplication and addition processes: leading to reduced latency. The lowest latency is achieved when n is power of two, while the worst result is given when n is a prime number or has large prime factors. Therefore, with changing set sizes, we observe a fluctuation in computational time. Of course, based on Figs. 5 and 6, this reduced computational overhead is not free and comes at a price. Nevertheless, we emphasize that this performance level is enough to support applications to detect the number of speakers in a target environment. More complex algorithms may be able to output other interesting results, but they are not of our interest.

Based on the observations from our MATLAB testing environment, we now move our focus to an empirical experiment setting, where we implement the proposed peak detection-based microphone array processing scheme (with time frame calibration) on our target hardware platform. Using this experimental prototype, we deploy the microphone in a room with a size of $\sim$60 m^2 and ask seven different volunteers to make natural vocal sounds at predefined locations while keeping track of the ground truth. A total of 20 different angles were tested with a total of 500 vocal activities, and each microphone of the array was sampled at 44 kHz for 250 ms. We present the estimation errors observed for each unique ground truth speaking angle in Fig. 8. As also validated in MATLAB, notice that the average DoA error is $\pm$ $\sim$1.5 degrees. This leads

Fig. 9. Picture of PIR installation for both entrance detection and idle state detection.

to two different conclusions. First, this is evidence that the microphone array sensor processing scheme works effectively under the constraints of resource-limited computing platforms. Furthermore, the average DoA error provides us with an intuition on how ϵ can be configured. Since the average speaker angle detection error is $\pm$ $\sim$1.5 degrees, configuring ϵ to be twice of this range (e.g., $\pm$ $\sim$3 degrees) will assure that the parameter is suitable for our system. Overall, these results indicate that despite the use of a lightweight classification scheme, the microphone array sensing platform in ReLiSCE holds the capability to well-separate the monolog and discussion states.

B. PIR Sensing Platform Evaluation

We now evaluate the performance of our PIR sensing platform in classifying the room-enter, room-idle, and room-in-use states. In doing so, we noticed that the installation position of the PIR sensor can significantly impact the detection performance. To validate the performance of the PIR sensing platform in various positions, we test the PIR sensor installed in two different locations. In the first, we install the PIR sensor to the ceiling of the entrance door, and in the second, the PIR sensor is installed to the ceiling above the discussion table, which is located at the center of the meeting room. We present pictures of these two locations in Fig. 9.

Using PIR sensors installed in these two positions, we first evaluate the detection performance of the room-enter state. We asked volunteers to perform natural meeting room entering actions, both alone and in groups. Furthermore, to test and compare the performance of using the microphone array sensor (e.g., installed at the center of the discussion table) for detecting the room-enter state, we ask the volunteers to make natural discussions when entering the room in groups. The results extracted from 60 room-entering actions (e.g., 30 single-person entrances and 30 group entrances) are presented in Fig. 10. We point out that ReLiSCE reports a "valid detection" only if the detection occurs within 1 s of the actual entering motion. Notice from the results that when the PIR sensor is attached to the ceiling of the entrance, the detection ratio of the room-enter state is $\sim$100%. On the other hand, when the sensor is placed in a more general location (e.g., ceiling of the center of meeting table), the detection ratio drops to $\sim$93%. This is due to the limited detection angle of our PIR sensor, which is $\sim$120°; suggesting that the quantity and locations of PIR sensors should be carefully determined based on application requirements. We can also notice that the detection ratio of the PIR sensor is higher than that of the microphone. The main cause of the microphone's lower detection ratio is

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

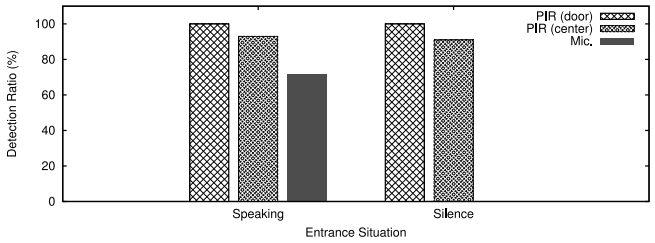


Fig. 10. Entering state detection with PIR and microphone array. We test for PIR sensors in two different locations and also test for the case where the participants enter the room as they make natural discussions and silently.

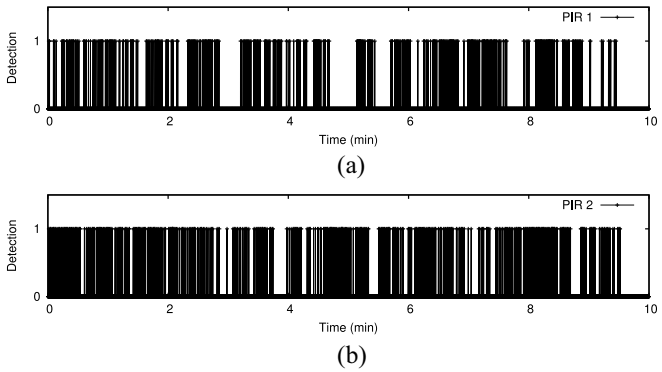


Fig. 11. Occupancy detection in a multiperson discussion scenario for PIR sensors installed at two different locations. (a) PIR sensor installed at room entrance. (b) PIR sensor installed on top of discussion table.

the fact that it is located far from the entrance: on the center of the discussion table. Therefore, small vocal sounds made near the door are considered as background noise and may not be properly detected. This difference in performance between the two sensing modalities implicitly shows the importance of using PIR sensors when detecting the room-enter state. Nevertheless, depending on the location of the PIR sensor, the microphone can still act as a supplementary sensor for detecting human entrance (or room occupancy).

While the PIR on the ceiling of the discussion table shows relatively lower entrance detection ratio, the location of this sensor makes it more suitable to continuously monitor the occupancy state of the meeting room. During a typical meeting, participants move their hands, change postures, or perform note taking actions. While it is difficult to distinguish between these small personal-scale actions, ReLiSCE exploits the PIR sensor to extract the fact that the room is still occupied. The question is how sensitive these sensors are to people's movements, and how they effectively detect occupancy under real meeting scenarios. We experimentally measure this performance using the two PIR sensors installed above and capture traces from typical meeting scenarios. Specifically, we captured 15 real meetings, resulting in an ∼18 h of trace. In Figs. 11 and 12, we present a subset of this duration while noting that other time instances showed similar behavior. Specifically, in Fig. 11, we present PIR detection traces for our two PIR sensors when more than two people perform a discussion-based meeting. Notice that both sensors actively detect movements of the participants, indicating that the occupancy during typical meetings can be properly detected. Next, in Fig. 12, we present results for a case where

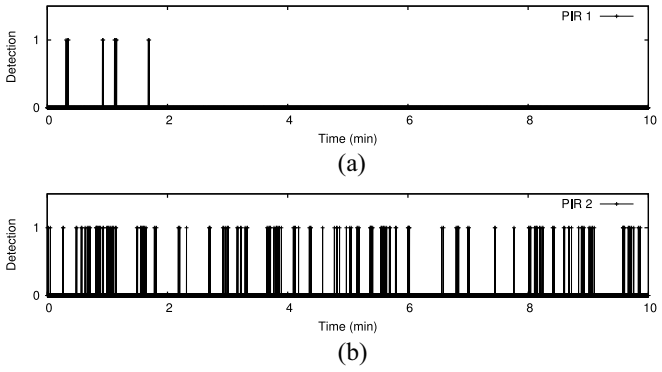


Fig. 12. Occupancy detection in a single person, solo-work scenario for PIR sensors installed at two different locations. (a) PIR sensor installed at room entrance. (b) PIR sensor installed on top of discussion table.

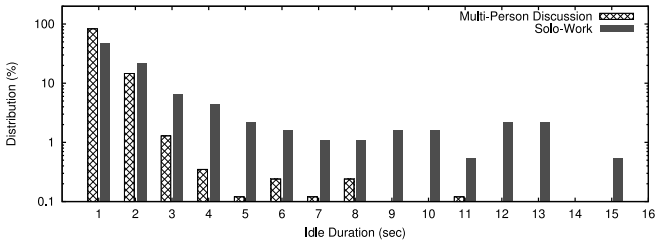


Fig. 13. Idle duration of array PIR distribution for different cases, discussing and sitting on the table, during 10 min.

TABLE I
MEAN, STANDARD DEVIATION, AND MAXIMUM IDLE DURATION OF PIR SENSORS FOR DIFFERENT CASES, MULTIPERSON DISCUSSION, AND SINGLE-PERSON SOLO-WORK

Status	Mean Duration	Std.Dev	Max Idle Duration
Discussion	0.66 sec	0.74	10.10 sec
Solo-Work	3.25 sec	5.43	32.4 sec

there is no active discussion, and a single person enters the meeting room to complete his or her task with a laptop computer. This activity involves mostly keyboard typing, wireless mouse-based input, and tilting one's head to observe the laptop screen. Fig. 12 suggests that despite the reduced activity level (compared to a multiperson discussion action), the PIR sensor positioned above the meeting room table is sensitive enough to capture human activity and identify the room-in-use state. Again, here, we can see that the installation positions of the PIR sensors can give significantly different detection results.

Finally, to evaluate the detection performance of the room-idle state, we analyze the collected results and plot the distribution of intermediate idle durations within the room-in-use state. We plot the results in Fig. 13 and summarize them as a table in Table I. Note that while room entrance can be detected instantaneously, detecting the room-idle state requires a level of delay for the system to confirm that there is no longer activity in the room. Our results provide data on such wait durations collected from real meetings. We can notice from Fig. 13 that when our volunteers were discussing, ∼99% of the idle durations were $\leq$ 8 s. On the other hand, since a "solo-work" scenario shows less human movement and interaction, ∼99% of the idle detections were within 14 s.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

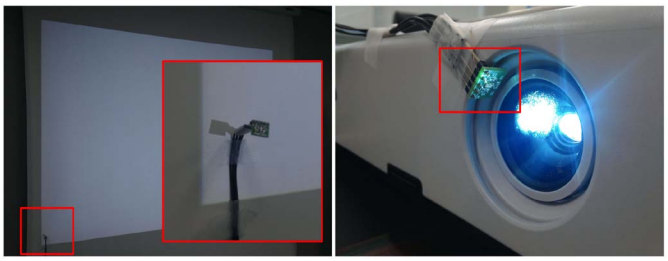


Fig. 14. Picture of illumination sensor on the projection screen and on the light bulb of the projector.

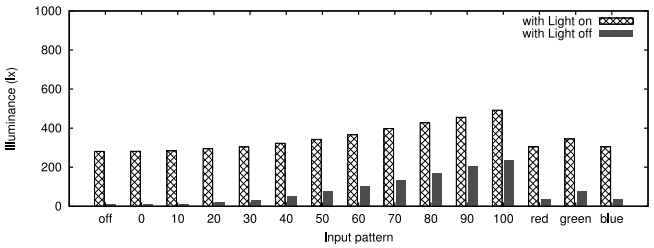


Fig. 15. Light sensor readings, when installed to the projection board, while projecting different RGB colors and gray scale (normalized to 0–100) changes both with and without ambient light.

This suggests that if the room is detected to be idle by the PIR sensing platform for $\geq$ 15 s, it is safe for the system to conclude that the room is idle. If this duration is shorter, there is a possibility that the ReLiSCE detects the wrong human activity context (e.g., triggers room-idle state while people are present in the room). While it is true that different meetings may show other numbers, we believe that our data set covers diverse types of meetings and therefore, these numbers can be used as a representative threshold. Nevertheless, we emphasize that the main contribution of this evaluation is not in identifying the exact threshold for room-idle state detection, but rather the contribution is in experimentally validating the feasibility and effectiveness of using PIR sensors for monitoring the occupancy level with various real-world intrameeting room activities.

C. Illumination Sensing Platform Evaluation

Lastly, we evaluate the performance of ReLiSCEs illumination sensing unit to capture the activity of beam projectors, which are widely used in presentation-oriented meetings. Similar to the PIR sensor, we determined that the installation position will significantly impact the sensor's detection performance; thus, we test for cases where the illumination sensor is positioned in two different locations as pictured in Fig. 14. Specifically, first, we test a scenario where the illumination sensor is attached to the projection screen, and second, we position the illumination sensor next to the light-bulb of the projector. In this evaluation, we focus on confirming that the illumination sensor can detect the projector's activity with high accuracy. For this, we first take a look at the detected/measured illumination levels at our sensing component with respect to different projection colors and the effect of ambient (fluorescent) light conditions.

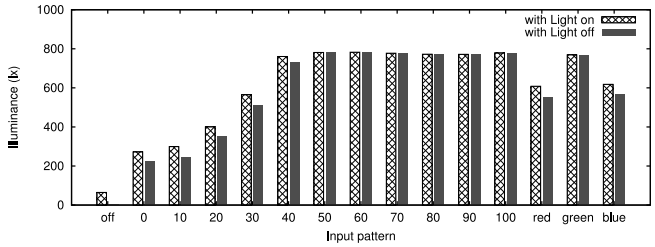


Fig. 16. Light sensor readings, when installed to the projector's light bulb, while projecting different RGB colors and gray scale (normalized to 0–100) changes both with and without ambient light.

First, in Fig. 15, we present the illumination level observed when the sensor was installed at the corner of the projection screen. Notice from the results that ambient lighting conditions have heavy impact on the illumination levels detected at our sensor. In this case, the illumination sensor was facing toward the projection unit; thus, the room lighting conditions directly affect its readings. As a result, through Fig. 15, we can notice that it is difficult to come up with a single threshold that distinguishes the projector's on/off states under any of the gray-scale intensity and color outputs. Therefore, despite having accurate microphone array readings of monolog activity, having the illumination sensor attached to the projection screen makes it difficult for ReLiSCE to properly determine the presentation state.

We next try attaching the illumination sensor adjacent to the beam projector's light-bulb and present the detected illumination levels in Fig. 16. Since the illumination sensor is closer to the light source, and also is facing downward toward the light-bulb (rather than the direction of ambient light sources), the results are only minimally affected by external lighting conditions. Furthermore, results in Fig. 16 also suggest that in this scenario, the sampled values of the illumination sensor is distinguishable despite displaying projection patterns with low gray scale density. Therefore, it becomes possible to configure a threshold that distinguishes the beam projector's activity. Based on these results, in ReLiSCE, we set a threshold of $\sim$200 lux. From a trial of 20 real use cases of the projection unit (average usage duration of 40 min per use), we were able to notice that this threshold allowed continuous detection of the projection state throughout the meeting durations.

The distinctive measurement patterns of the illumination sensor for detecting the projector's activity state, and the accurate angular detection of speakers' DoA using our microphone array sensor suggests that the presentation state can be detected with very high accuracy. In fact, our empirical experiences showed a 95% proper detection ratio of the presentation state. We note that for the 5% of the misclassified cases, external vocal sounds caused the microphone array sensor to falsely detect the discussion state (e.g., rather than the monolog state used for determining the presentation state).

D. Discussions on Connecting ReLiSCE With Applications

Overall, our evaluations focus on quantifying the performance of ReLiSCEs sensing components in various real meeting environments. Through these tests, we were able to notice that the individual sensing components' performance

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

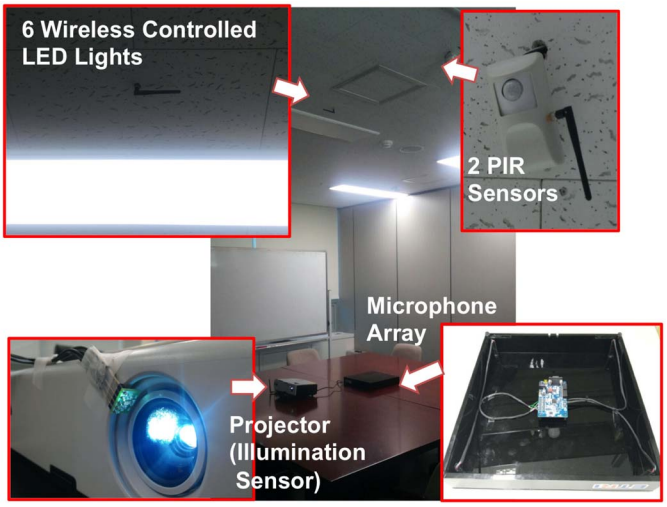


Fig. 17. Picture of smart lighting system with wirelessly controlled LED lighting components.

can be significantly improved with careful node deployment. Thanks to the satisfactory performance of each component, we've noticed that human activity states can be estimated with high reliability using heterogeneous sensing modalities. Given such a system, the next step is to combine the extracted activity states from ReLiSCE with real application systems.

One example of an application that can greatly benefit from ReLiSCE is a smart lighting application. To validate the effectiveness of ReLiSCE in such a system, we installed wireless modules to the lighting components in our meeting room environment (see Fig. 17) and used the results from ReLiSCE for controlling the lighting conditions. The algorithms for controlling the lighting levels were designed based on an extensive set of feedback from domain experts. Nevertheless, based on a ∼45 day test trial, the user satisfactory level of the ReLiSCE-based smart lighting system (based on manual surveys from users) were at 93%. The 7% of the negative feedback was mostly caused by wrong decision making from ReLiSCE, and also by the sensitive actuation made at the lighting components. Specifically, most of these false detections were caused due to the sensitivity of our microphone array. We plan to address drawback on both hardware and software perspectives as part of our future work.

V. Related Work

With advances in sensing devices, research on smart office activity context extraction systems has been widely studied over the past few decades [10], [11]. These systems were mostly dependent on sound- or vision-based sensors for capturing the human activities [12], [13]. While vision-based approaches have great potentials for extracting sophisticated activity contexts, they require excessive computational operations for image processing; thus, are not suitable for resource-limited embedded computing platforms. Furthermore, image-based systems introduce privacy threats to the people residing in the environment, making them more suitable for intrusion detection applications than activity extraction. In contrast, for a computational environment such as ours, sound-based systems are more suitable, which can be generally classified into two major paradigms: sound source localization and classification.

Sound source localization systems are designed to locate the sources of the vocal sounds for counting the number of speakers in the target environment [7], [14]. For this purpose, as in ReLiSCE, a microphone array is widely used [15]. To localize the sound source, these systems mostly utilize one of the three following methods.

1) Steered beamforming identifies the sound source through a procedure called focalization [16], which rotates the microphone array to search for the direction that yields the strongest signal strength. Although efforts have been put into increasing the performance, this method is not yet practical to use due to its high computational complexity, and sensitivity to background noise [17].
2) High-resolution spectral estimation is an approach applied to wide-band signals (e.g., human speech) [18], [19]. Here, sound signals received at each microphone is used to derive a spatio-spectral correlation matrix for speaker location estimation. Nevertheless, this approach also introduces a high computational overhead and holds a strong assumption that the speaking object is always statistically stationary [20].
3) TDOA-based estimation, also used in ReLiSCE, relies on relative time delays between received signals rather than the characteristics of the acoustic signals. In practice, TDOA is estimated by identifying the time shift that maximizes the correlation between a pair of sound signals [21], [22]. Among various methods, as previously mentioned and compared, GCC-PHAT [23] is known to provide a reliable performance [24], [25]. Despite its attractiveness in noisy environments, GCC-PHAT is computationally hungry; making it difficult to implement on resource-limited platforms [7], [26].

Unlike sound source localization, which focuses on inferring the activity contexts with respect to speakers' location, sound source classification extracts activity context based on the type of sound. This method is applicable when a system is looking for a specific type of sound (e.g., speech, footsteps, or door closing sounds) [26]–[29]. For meeting room context extraction, analyzing, and classifying speech can provide quality features for automating meeting content indexing systems [11], [30] by adapting the group task model [4], [5], [31]. However, again, these classification methods require a heavy level of computation (e.g., feature extraction and comparison process), and they are limited to a limited number of preknown sounds. Due to such limitations, ReLiSCE does not focus on sound classification.

We note that some systems use only the microphone sensor for indoors activity context extraction [28]. Nevertheless, the reduced number of sensing components naturally leads to increased complexity on the processing software. To reduce this overhead, ReLiSCE introduces an additional set of heterogeneous sensors to supplement the decisions made at the microphone array sensor.

Finally, while there is no existing system yet in the literature, it is arguable that widely deployed smartphones can

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

be applicable for application scenarios that ReLiSCE targets. Specifically, we could eventually utilize the microphone units and the high processing power of smartphones to perform the speaker detection processes of the microphone array sensor in ReLiSCE using traditional GCC-PHAT-based methods. However, we find this approach not yet feasible due to two major reasons. First, it is currently difficult to allow smartphones positioned *ad hoc* to accurately detect the speaker's direction with low angular errors. This could eventually be made possible using range estimation techniques such as BeepBeep [32], but such techniques are not yet practical to use in general scenarios due to several requirements such as the need of line of sight. Even if the distance and locations of these smartphones were known, the frequent use of smartphones (even during meetings) make the estimation process even more challenging on a practical perspective. Second, continuous acoustic signal processing on smartphones will require them to use a significant amount of their computational power. Such continuous background processing can significantly degrade the user experience and lead to lower utilization in the services it provides.

VI. Summary

This paper started with a simple question in our mind: "can we compress the capabilities of a microphone sensor to fit in resource-limited platforms and capture the human activity states?" Our proposed system, ReLiSCE, reveals our answer: not alone, but together, yes. The limited capability of a single embedded computing platform results in the microphone array sensor to only extract monolog and discussion states, but the help of a heterogeneous set of sensors, such as illumination and PIR sensors, allows ReLiSCE to accurately determine various human activity states in an office meeting room environment with near-perfect accuracy. As briefly presented earlier, we believe that ReLiSCE should be tightly integrated with application systems, such as smart lighting applications, to further increase the human comfort levels and maximize the operational efficiency (e.g., energy efficiency). Furthermore, we believe that systems such as ReLiSCE can open the possibilities of quickly commercializing human activity context extraction systems for everyday users by providing them with an easy-to-install system that can be applied to various application scenarios. Lastly, we foresee that the decentralization of complex computational operations for use in distributed embedded computing platforms can introduce interesting questions as we finally try to realize various internet of things applications.

References

[1] C. Lin, C. Federspiel, and D. Auslander, "Multi-sensor single actuator control of HVAC systems," in *Proc. Int. Conf. Enhanc. Build. Oper.*, Austin, TX, USA, 2002.

[2] A. Schaeper, C. Palazuelos, D. Denteneer, and O. Garcia-Morchon, "Intelligent lighting control using sensor networks," in *Proc. 10th IEEE Int. Conf. Netw. Sens. Control (ICNSC)*, Evry, France, Apr. 2013, pp. 170–175.

[3] J. Ko *et al.*, "MEDiSN: Medical emergency detection in sensor networks," *ACM Trans. Embedded Comput. Syst.*, vol. 10, no. 1, Aug. 2010, Art. ID. 11.

[4] J. E. McGrath, *Groups: Interaction and Performance*, vol. 14. Englewood Cliffs, NJ, USA: Prentice-Hall, 1984.

[5] I. McCowan *et al.*, "Automatic analysis of multimodal group actions in meetings," *IEEE Trans. Pattern Anal. Mach. Intell.*, vol. 27, no. 3, pp. 305–317, Mar. 2005.

[6] CRZ Technology. (2014). *Mango Z1 Board Specifications*. [Online]. Available: http://www.mangoboard.com/sub2.html?ptype=view&prdcode=1105170004

[7] D. Sun and J. Canny, "A high accuracy, low-latency, scalable microphone-array system for conversation analysis," in *Proc. 2012 ACM Conf. Ubiquitous Comput. (UbiComp)*, Pittsburgh, PA, USA, pp. 290–300.

[8] "BH1600FVC—Analog current output type ambient light sensor IC BH1600FVC," Rohm Co. Ltd., Kyoto, Japan, Tech. Note 12046EDT04, 2012.

[9] D. Moss, J. Hui, and K. Klues, "Low power listening," *TinyOS Core Working Group TEP*, vol. 105, 2007.

[10] S. Renals, T. Hain, and H. Bourlard, "Recognition and understanding of meetings the AMI and AMIDA projects," in *Proc. 2007 IEEE Workshop Autom. Speech Recognit. Und. (ASRU)*, Kyoto, Japan, pp. 238–247.

[11] Z. Yu and Y. Nakamura, "Smart meeting systems: A survey of state-of-the-art and open issues," *ACM Comput. Surv.*, vol. 42, no. 2, pp. 8:1–8:20, 2010.

[12] A. Janin *et al.*, "The ICSI meeting project: Resources and research," in *Proc. 2004 IEEE Int. Conf. Acoust. Speech Signal Process. Meeting Recognit. Workshop (NIST RT)*, Montreal, QC, Canada.

[13] V. Stanford, J. Garofolo, O. Galibert, M. Michel, and C. Laprun, "The NIST smart space and meeting room projects: Signals, acquisition annotation, and metrics," in *Proc. 2003 IEEE Int. Conf. Acoust. Speech Signal Process. (ICASSP)*, vol. 4. Hong Kong, pp. 6–10.

[14] A. Willig and R. Mitschke, "Results of bit error measurements with sensor nodes and casuistic consequences for design of energy-efficient error control schemes, in *Wireless Sensor Networks*, Berlin, Germany: Springer, 2006, pp. 310–325.

[15] X. Bian, G. D. Abowd, and J. M. Rehg, "Using sound source localization in a home environment," *in Proc. 3rd Int. Conf. Pervasive Comput.*, Munich, Germany, 2005, pp. 19–36.

[16] W. J. Bangs and P. M. Schultheiss, "Space-time processing for optimal parameter estimation," in *Signal Processing*. New York, NY, USA: Academic, 1973, pp. 577–590.

[17] H. F. Silverman and S. E. Kirtman, "A two-stage algorithm for determining talker location from linear microphone array data," *Comput. Speech Lang.*, vol. 6, no. 2, pp. 129–152, 1992.

[18] R. O. Schmidt, "Multiple emitter location and signal parameter estimation," *IEEE Trans. Antennas Propag.*, vol. 34, no. 3, pp. 276–280, Mar. 1986.

[19] H. Wang and M. Kaveh, "Coherent signal-subspace processing for the detection and estimation of angles of arrival of multiple wide-band sources," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 33, no. 4, pp. 823–831, Aug. 1985.

[20] M. S. Brandstein and H. F. Silverman, "A practical methodology for speech source localization with microphone arrays," *Comput. Speech Lang.*, vol. 11, no. 2, pp. 91–126, 1997.

[21] A. Lombard, Y. Zheng, H. Buchner, and W. Kellermann, "TDOA estimation for multiple sound sources in noisy and reverberant environments using broadband independent component analysis," *IEEE Audio, Speech, Language Process.*, vol. 19, no. 6, pp. 1490–1503, Aug. 2011.

[22] X. Sheng and Y. H. Hu, "Maximum likelihood multiple-source localization using acoustic energy measurements with wireless sensor networks," *IEEE Trans. Signal Process.*, vol. 53, no. 1, pp. 44–53, Jan. 2005.

[23] C. Knapp and G. C. Carter, "The generalized correlation method for estimation of time delay," *IEEE Trans. Acoust. Speech Signal Process.*, vol. 24, no. 4, pp. 320–327, Aug. 1976.

[24] S. Araki, R. Mukai, S. Makino, T. Nishikawa, and H. Saruwatari, "The fundamental limitation of frequency domain blind source separation for convolutive mixtures of speech," *IEEE Audio, Speech, Language Process.*, vol. 11, no. 2, pp. 109–116, Mar. 2003.

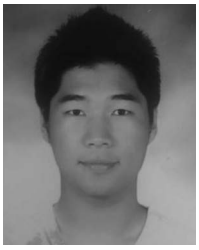
[25] M. Wu and D. Wang, "A two-stage algorithm for one-microphone reverberant speech enhancement," *IEEE Audio, Speech, Language Process.*, vol. 14, no. 3, pp. 774–784, May 2006.

[26] L. Gu *et al.*, "Lightweight detection and classification for wireless sensor networks in realistic environments," in *Proc. ACM 3rd Int. Conf. Embedded Netw. Sensor Syst. (SenSys)*, San Diego, CA, USA, 2005, pp. 205–217.

[27] F. Bimbot *et al.*, "A tutorial on text-independent speaker verification," *EURASIP J. Appl. Signal Process.*, vol. 2004, pp. 430–451, Jan. 2004.

This article has been accepted for inclusion in a future issue of this journal. Content is final as presented, with the exception of pagination.

[28] P. Delacourt and C. Wellekens, "DISTBIC: A speaker-based segmentation for audio data indexing," *Speech Commun.*, vol. 32, no. 12, pp. 111–126, 2000.
[29] Y. Guo and M. Hazas, "Localising speech, footsteps and other sounds using resource-constrained devices," in *Proc. 2011 10th Int. Conf. Inf. Proc. Sensor Netw. (IPSN)*, Chicago, IL, USA, pp. 330–341.
[30] D. Gatica-Perez, "Automatic nonverbal analysis of social interaction in small groups: A review," *Image Vis. Comput.*, vol. 27, no. 12, pp. 1775–1787, 2009.
[31] D. Zhang, D. Gatica-Perez, S. Bengio, and I. McCowan, "Modeling individual and group actions in meetings with layered HMMs," *IEEE Trans. Multimedia*, vol. 8, no. 3, pp. 509–520, Jun. 2006.
[32] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, "BeepBeep: A high accuracy acoustic ranging system using COTS mobile devices," in *Proc. ACM 5th Int. Conf. Embedded Netw. Sensor Syst. (SenSys)*, Sydney, NSW, Australia, 2007, pp. 1–14.



Homin Park received the B.S. degree in computer science and systems from the University of Washington, Tacoma, WA, USA. He is currently pursuing the Ph.D. degree in the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea.

His current research interests include cyber-physical systems, context extraction systems, smart home environments, wireless sensor networks, and intelligent vehicular safety systems.

Jongjun Park received the B.S. and M.S. degrees in electrical engineering from the Pohang University of Science and Technology, Pohang, Korea, in 2004 and 2006, respectively.

He is currently a Senior Researcher with the Electronics and Telecommunications Research Institute, Daejeon, Korea. His current research interests include designing wireless systems for smart environments and analyzing wireless signal characteristics.

Hyunhak Kim received the B.S. degree from Kyungpook National University, Daegu, Korea, in 2004, and the M.S. degree in computer science from the Korea Advanced Institute of Science and Technology, Daejeon, Korea.

He is currently with the Electronics and Telecommunications Research Institute, Daejeon.

Jongarm Jun received the bachelor's and master's degrees, both in electrical engineering, from Kyungbuk National University, Daegu, Korea, and Yonsei University, Seoul, Korea, in 1987 and 1989, respectively.

Since 1989, he has been a Principal Researcher at the Electronics and Telecommunications Research Institute, Daejeon, Korea. His current research interests include designing systems for the Internet of Things.

Sang Hyuk Son (F'13) received the B.S. degree in electronics engineering from Seoul National University, Seoul, Korea, the M.S. degree from the Korea Advanced Institute of Science and Technology (KAIST), Daejeon, Korea, and the Ph.D. degree in computer science from the University of Maryland, College Park, College Park, MD, USA.

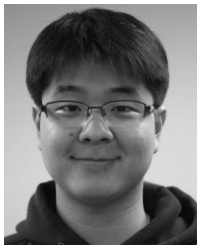
He is a Department Chair of Information and Communication Engineering with the Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea. He has been a Professor in the Computer Science Department, University of Virginia, Charlottesville, VA, USA, and a World Class University Chair Professor at Sogang University, Seoul, Korea. He has been a Visiting Professor at KAIST, City University of Hong Kong, Hong Kong, Ecole Centrale de Lille, Villeneuve-d'Ascq, France, Linköping University, Linköping, Sweden, and the University of Skövde, Skövde Sweden. His current research interests include cyberphysical systems, real-time and embedded systems, database and data services, and wireless sensor networks. He has authored or coauthored over 290 papers and edited/authored four books in the above areas.

Prof. Son was the recipient of the Outstanding Contribution Award from ACM/IEEE Cyber Physical Systems Week in 2012. He has served as a Chair of the IEEE Technical Committee on Real-Time Systems from 2007 to 2008. He is serving as an Associate Editor of the *Real-Time Systems Journal* and the *Journal of Computing Science and Engineering*, and has also served on the editorial board of the IEEE TRANSACTIONS ON COMPUTERS and the IEEE TRANSACTIONS ON PARALLEL AND DISTRIBUTED SYSTEMS. He is also serving as a Steering Committee Member of RTCSA, Cyber Physical Systems Week, and IFIP Workshop on Software Technologies for Future Embedded and Ubiquitous Systems. His research has been funded by the National Science Foundation, Defense Advanced Research Projects Agency, Office of Naval Research, Department of Energy, National Security Agency, IBM, and the National Research Foundation of Korea.

Taejoon Park (M'05) received the B.S. degree (*summa cum laude*) in electrical engineering from Hongik University, Seoul, Korea, the M.S. degree in electrical engineering from the Korea Advanced Institute of Science and Technology, Daejeon, Korea, and the Ph.D. degree in electrical engineering and computer science from the University of Michigan, Ann Arbor, MI, USA, in 1992, 1994, and 2005, respectively.

He is an Associate Professor with the Department of Information and Communication Engineering, Daegu Gyeongbuk Institute of Science and Technology, Daegu, Korea. He was an Assistant Professor at Korea Aerospace University, Gyeonggi-do, Korea, from 2008 to 2011, a Principal Research Engineer at Samsung Electronics, Gyeonggi-do, from 2005 to 2008, and a Research Engineer at LG Electronics, Seoul, from 1994 to 2000, and was promoted to a Senior Research Engineer in 2000. His current research interests include cyberphysical and networked embedded systems with emphasis on smartness, reliability, and timeliness. He has authored or coauthored over 100 papers/patents including essential patents for the Digital Video Disk standard, five of which were cited over 100 times. He has an H-index of 14 with over 1000 Google Scholar citations.

Prof. Park is a member of the Association for Computing Machinery.



JeongGil Ko (M'06) received the B.Eng. degree in computer science and engineering from Korea University, Seoul, Korea, in 2007, and the M.S.E. and Ph.D. degrees in engineering and computer science, both from the Johns Hopkins University, Baltimore, MD, USA, in 2009 and 2012, respectively.

Since 2012, he has been with the Internet of Things (IoT) Convergence Research Department at the Electronics and Telecommunications Research Institute, Daejeon, Korea, as a Researcher. In 2010, he was a Visiting Researcher at the Stanford Information Networking Group with Dr. P. Levis at Stanford University, Stanford, CA, USA. His current research interests include the general area of developing web and cloud-based sensing systems with ambient intelligence for the IoT and cyberphysical systems.

Dr. Ko was the recipient of the Abel Wolman Fellowship awarded by the Whiting School of Engineering at the Johns Hopkins University, Baltimore, MD, USA, in 2007. He was a member of the Hopkins interNetworking Research Group (HiNRG) led by Dr. A. Terzis.