2015 IEEE International Conference on
Automation Science and Engineering (CASE)
Aug 24-28, 2015. Gothenburg, Sweden

SoundLoc: Accurate Room-level Indoor Localization using Acoustic Signatures

Ruoxi Jia, Ming Jin, Zilong Chen and Costas J. Spanos

Abstract— **Room-level indoor localization is of particular interest in the energy-efficient smart building, as services, such as lighting and ventilation, can be targeted towards individual rooms based on occupancy instead of an entire floor. Hence, this paper focuses on identifying the room where a person or a mobile device is physically present. Existing room-level localization methods, however, require special infrastructure to annotate rooms with special signatures. SoundLoc is a room-level localization scheme that exploits the intrinsic acoustic properties of individual rooms and obviates the needs for infrastructures. As we will show in the study, rooms' acoustic properties can be characterized by Room Impulse Response (RIR). Nevertheless, obtaining precise RIRs is a time-consuming and expensive process. The main contributions of our work are the following: First, a cost-effective RIR measurement system is designed and the Noise Adaptive Extraction of Reverberation (NAER) algorithm is developed to estimate room acoustic parameters in noisy conditions. Second, a comprehensive physical and statistical analysis of features extracted from RIRs is performed. Also, SoundLoc is evaluated using the dataset consisting of ten (10) different rooms and the overall accuracy of 97.8% has been achieved.**

I. Introduction

Commercial buildings contribute to 19% of the primary energy consumption in US. Prior research has shown that most of the buildings use static control for building facilities, such as Heating, Ventilation and Air Conditioning (HVAC) and lighting systems, and thereby considerable energy is wasted in conditioning and lighting unoccupied spaces. Awareness of occupancy information can help adaptively run the conditioning and lighting systems and reduce energy consumption in buildings [1], [2]. Therefore, the availability of occupants' indoor positions has become an immediate need.

Unlike outdoors, where the Global Positioning System (GPS) can provide a relatively accurate and robust solution for positioning, indoor localization has not been equally facilitated by GPS due to significant positioning error of satellite-based navigation systems in closed environments. A variety of alternative methods have been proposed, ranging from visual [3] to infrared [4]. There has also been extensive research work focusing on indoor localization systems based on WiFi wireless network along with WiFi enabled devices [5]. However, the density of access points has a strong influence on localization accuracy. The reported WiFi localization accuracy falls below 70% in real usage environments since access point density may be low or occupancy variation may lead to significant fluctuations in WiFi signals. Also, these techniques have certain disadvantages that special-purpose infrastructures are required to support localization.

In contrast to infrastructure-based localization schemes, SoundLoc takes advantage of internal microphones and speakers on laptops or mobile phones, the most ubiquitous devices, to measure the acoustic properties of the space where people are physically present. No specialized infrastructure is required to be pre-installed for localization. A key observation that supports our work is that the indoor environment is well-structured and can be organized into spaces with distinct geometries and functionalities. These spaces can either be open and without proper boundaries such as hallways, or closed such as offices. Herein, we use "room" in a broad sense to refer to both open and closed spaces. We notice that the control of lighting and HVAC systems are typically applied at the granularity of room, and therefore a room-level localization is sufficient for any occupancy-aware control of lighting and HVAC systems. Radio-based techniques may easily confuse nearby rooms as the random variation of radio signals may induce a poor distance estimate and thereby place the person in the incorrect adjacent rooms. However, our work identifies rooms by exploiting rooms' intrinsic acoustic effects that are governed by the geometry and furnishings. Even though two rooms are geospatially adjacent, they can be easily distinguished in the acoustic feature space.

SoundLoc offers an accurate solution for room-level indoor localization using acoustic signatures of rooms and requires no specialized infrastructure to support localization. The rest of this paper is organized as follows. Section II describes related work. Section III formulates the room-level localization problem in terms of rooms' acoustic properties. Section IV describes the experimental design. Section V explores various acoustic features that are promising to be used for localization. Section VI evaluates the performance of SoundLoc. Section VII concludes the paper.

II. Related Work

Extensive work on indoor localization has been focusing on creating unique fingerprints for specific positions. Various types of fingerprints have been developed, mainly

*This research is funded by the Republic of Singapores National Research Foundation through a grant to the Berkeley Education Alliance for Research in Singapore (BEARS) for the Singapore-Berkeley Building Efficiency and Sustainability in the Tropics (SinBerBEST) Program. BEARS has been established by the University of California, Berkeley as a center for intellectual excellence in research and education in Singapore.

R. Jia, M. Jin and C.J. Spanos are with the Department of Electrical Engineering and Computer Sciences at the University of California Berkeley, USA. ruoxijia@berkeley.edu, jinming@berkeley.edu, spanos@berkeley.edu

Z. Chen is with the Department of Electronic Engineering at Tsinghua University, China. chenzl11@mails.tsinghua.edu.cn

RF and ambient fingerprints. Radar [6] pioneered fingerprinting method based on received signal strength hearing from multiple access points. ARIEL [7] proposed a room localization system that correlates WiFi signal strength with occupants motion patterns to improve accuracy. However, these techniques are hampered by long-term signal variations and low access point density. SurroundSense [8] extends the fingerprinting idea by fusing multiple sensor measurements collected on smartphone and infers the position from various ambient features such as background sound level, light, color as wll as WiFi signal strength. In addition to sound amplitude, the spectral features of sound are also applied to fingerprint the space. For instance, ABS [9] identifies rooms by exploiting the spectral characteristics of background sound, while SoundSense [10] analyzes the transient sound existing in the environment. These works are based upon the assumption that the ambient sound features of the place of interest are stationary and informative, which might not always be satisfied. In reality, background sound can vary both slowly and transiently. For instance, people's talking could appear randomly and even HVAC's on/off states could generate distinctive background sounds that influence ambient sound fingerprints.

Rather than detecting the uncontrolled background sound, acoustic fingerprints, which characterizes the rooms acoustic effect on audio signals, have been proposed to fingerprint rooms. Peters et al. [11] demonstrate a system that identifies the room via analyzing acoustic features extracted from audio recordings. The accuracy of 61% has been achieved for musical signals and 85% for speech signals. Shabtai et al. [12] classifies the room based on reverberation time extracted from RIRs and achieves the error rate of 3.9%. However, the RIR samples used in the paper are collected from places that vary significantly in volumes and inside furnishings, such as classrooms, music hall, etc.

Our work differs from the aforementioned work from the following aspects. First, no extra specialized microphones need to be installed as we utilize internal microphone and speaker of mobile devices. Second, we leverage the rooms' intrinsic acoustic properties rather than analyzing the non-stationary background sound. Third, instead of using RIR samples available online that were collected from the spaces varying considerably in volumes, we develop a cost-effective RIR measurement system and collect RIR samples in several similar indoor environments, such as adjacent offices. We further demonstrate the effectiveness of acoustic features in identifying rooms in buildings, including both closed and open spaces.

III. Problem Formulation

For a sound signal generated inside a room, the sound may travel via the direct path from the source to the receiver or bounce off walls and other objects. Therefore, the received signal is a superposition of multiple delay and distorted versions of the original signal, perceived as echo and reverberation. Intuitively, the received signal contains information about rooms size and absorption properties. Since the indoor geometry and furnishings are roughly constant, we can approximate this "room effect" as a linear time-invariant system characterized by the impulse response $h(t)$. The received signal is the convolution of the transmitted signal and the room impulse response (RIR) in the time domain, as illustrated in Fig. 1. Since there exists a one-to-one mapping from a room to its room effect, an unknown room can be uniquely labeled theoretically if its RIR is available.

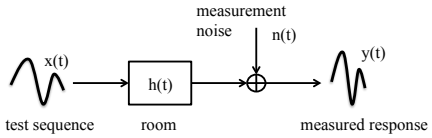

Fig. 1. The room can be modeled as a linear time-invariant system and the received signal is linear convolution of test sequence and room impulse response.

A common approach to measure the RIR is to apply a known excitation signal and measure the system's output and then deconvolve the measured response with the excitation signal. The choices concerning the excitation signals and deconvolution techniques are of essential significance to RIR measurement. Several types of most commonly used excitation signals are presented and compared in [13]. In our paper, we utilize Maximum Length Sequence (MLS) as the input sequence, which is known for its capability of providing superior dynamic range and high signal-to-noise ratio. MLS is a periodic pseudo-random signal and behaves similar to white noise stochastically. Hence, we can obtain the RIR by computing the autocorrelation of the received signal,

$$h(k) = R_y(k) = E[y(n)y(n-k)] \tag{1}$$

where $R_y(k)$ denotes the autocorrelation and $E[\cdot]$ is the expectation operator. In order to reduce the time-aliasing error, a MLS with longer period is preferable [14]. In our measurement system, the length of MLS is $2^{17}-1$. And we calculate autocorrelation using the fast Hadamard transform in order to reduce computational overhead [15]. As RIRs are essentially time series data and cannot be fed into the classification algorithm directly, it is necessary to extract "valuable" features from RIRs and these features should contain rich location information. We will explore different acoustic features in Section V.

IV. Experimental Configuration

The aim of our experiments is to verify if the noisy RIRs obtained by the cheap internal speakers and microphones on laptops contain features that are capable of indicating indoor locations. We also design experiments to further test the noise-robustness and time-invariance of the features.

A. Corpus Collection

We implement the MLS-based RIR measurement on laptops. The built-in loudspeaker plays a MLS sequence and the microphone records the sound signal simultaneously.

The whole playing and recording process last about 18 seconds. A fast devolution algorithm is running on laptop to compute RIRs and a 2.8 second CSV file (16-bit 44kHz) recording the RIR is generated. We extract various acoustic features that are potentially useful for localization from RIRs. Next, we deploy an extra feature selection step to determine "optimal" features in terms of distinctiveness, noise-robustness and time-invariance. Finally, different classifiers are experimented to determine the room label. Fig. 2 depicts a pictorial overview of SoundLoc's architecture.

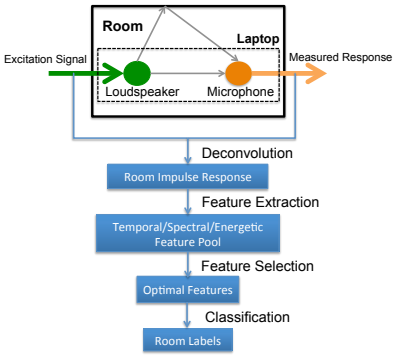

Fig. 2. SoundLoc Architecture

We collected RIR samples in 10 different functional areas in Cory Hall locating at the campus of UC Berkeley, the details of which are provided in Table I. The 10 areas include both closed and open spaces. These areas also vary in volumes, wall materials and furnishings. A description of their environments during data collection is also given in Table I. All of the 10 areas are controlled by different lightings, which allows the lighting system to turn off individual zone to save energy. The experimenter collect samples at two sub-locations in each room, 50 samples for each sub-location. All the experiments are carried out on ordinary workdays. Hence, the majority of our samples are collected with random background noise, such as speaking, footsteps, HVAC sounds, etc.

TABLE I
DETAILS OF INVESTIGATED ROOMS

Investigated Rooms	Area(m^2)	Description
Office A	10.2	Closed space, quiet
Office B	8.8	Closed space, quiet
Office C	7.1	Closed space, quiet
Office D	11.7	Closed space, quiet
Conference Room	26.1	Closed space, quiet
Lab	11.7	Closed space, server noise
Kitchen	6.5	Open space, speaking and coffee machine noise
Hallway	–	Open space, speaking noise
Stairs	–	Closed space, speaking, footstep and door creaky noise
Cubicle Zone	8.8	Open space surrounded by clipboard, speaking noise

B. Experimental Design

1) Experiment A: In this experiment, we aim at examining the distinctiveness of the features. We use the 1000 samples (10 rooms, 20 sub-locations in each room, 50 samples in each sub-location) in the way described above to construct our training sets and testing sets and conduct a 10-fold cross validation.

2) Experiment B: This experiment aims at examining the robustness of SoundLoc to noise. In particular, we collected 100 RIR samples in the conference room during and after a meeting. During the meeting, there exist successive talking and moving noise in the recordings. The testing set includes only noisy samples for conference room. The training set includes quiet samples for conference room and samples described in *Experiment A* for other places. Compared with *Experiment A*, this experiment is potentially challenging because the testing set is based on completely different samples used for training the model.

3) Experiment C: The purpose for this experiment is to test the time-stationarity of the features. We conduct the second visits in a different day to three areas: Office B, stairs and cubicle zone. These three locations are randomly picked from the places that are available. We do not exclude any samples after we see the result. For each place, 100 samples are collected. In the classification stage, the training set includes only the samples from the first visit while the test set includes only the samples from the second visit.

V. Acoustic Feature Exploration

A. Temporal Features

We use kurtosis of the RIR in time domain to capture its temporal properties. In statistics, kurtosis describes the peakedness of the probability density function of a real-valued random variable, given by

$$kur[h(k)] = \frac{E[(h(k)-\mu)^4]}{\sigma^4} \tag{2}$$

where μ is the mean of the RIR $h(k)$ and σ is the standard deviation. Higher kurtosis of a signal means more of the variance arises from infrequent extreme deviations, in contrast to frequent modestly sized deviations. The kurtosis of the RIR is an indicator of the volume of a room. If the room is large, the RIR will exhibit infrequent large deviations and thereby higher kurtosis, and vice versa. Fig. 3 illustrates the distributions of temporal kurtosis in different locations described in the Table I. As it can be seen, the closed spaces with relatively small volumes exhibit small kurtosis in the time domain, while the open spaces or spaces with large volumes have larger kurtosis.

B. Spectral Features

In acoustics, direct-to-reverberant energy ratio is an important parameter to characterize a room's acoustic properties. It depends on the geometry and absorption of the space where the sound waves propagate. In [**?**], it is shown that the standard deviation of RIR's spectrum increases with the direct-to-reverberant energy ratio and thereby spectral standard deviation can be used to characterize a room. Since

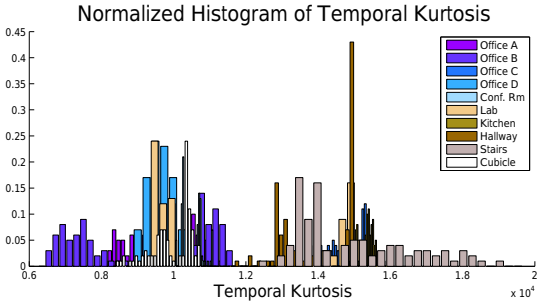

Fig. 3. Normalized histogram of temporal kurtosis. Closed spaces with small volumes such as offices, conference room and cubicle have smaller temporal kurtosis. Open spaces such as kitchen and hallway or closed spaces with large volume such as stairs have large kurtosis.

the absorption properties of materials are a function of frequency, we further inspect this feature within different frequency bands, as defined by

$$std_{[f_1,f_2]}[H(f)] = E_{[f_1,f_2]}[H^2(f)] - E^2_{[f_1,f_2]}[H(f)] \quad (3)$$

where $H(f)$ denotes the Fourier transform of the RIR and $E_{[f_1,f_2]}[\cdot]$ denotes taking empirical expectation over the frequency band ranging from f_1 to f_2. The distributions of spectral standard deviation in different rooms are plotted in Fig. 4. The rooms investigated here exhibit different absorption properties. For instance, offices, lab, cubicle, conference room and hallway are covered with the carpet, which is a good sound-absorbent material. In these areas, sound energy is absorbed before it bounces around the space and generates reverberation. In this case, direct sound energy from the emitter to the receiver will dominate, and thereby these locations have a higher direct-to-reverberant sound ratio, i.e., a larger spectral standard deviation. In contrast, locations without special sound reduction, such as stairs and kitchen, exhibit a relatively small spectral standard deviation.

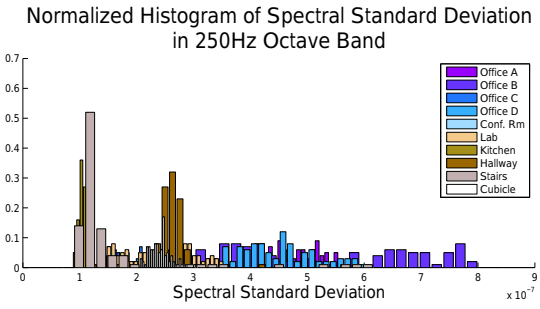

Fig. 4. Normalized histogram of spectral standard deviation in octave band centered at 250 Hz. Places covered with carpet such as offices, lab, hallway and cubicle exhibit a larger spectral standard deviation. Places without special sound reduction such as stairs and kitchen shows a smaller spectral standard deviation.

In addition, the kurtosis of Fourier coefficients is also included in our feature pool. This is because the room can be identified by its room modes which are collection of resonances that exist in a room when excited by a sound source and room modes can be noticed by magnitude peaks in the spectrum of the RIR. We use the kurtosis of Fourier coefficients to describe room modes and further to characterize a room.

C. Energetic Features

Energetic features describe how sound energy decays as it propagates in rooms. Reverberation Time (RT) is a promising energetic feature for room identification as it is insensitive to microphone arrangement and source orientation [**?**]. A standard RT is defined by the time taken for the acoustic energy in the space of interest to decay by 60 dB once the source is turned off. According to Sabines formula [**?**] in acoustic theory, $RT = 0.161\frac{V}{S\alpha}$, where RT is directly related to the volume V and the surface absorption of the room, which is the product of surface area S and average absorption coefficients α. RT can be estimated from the normalized energy decay curve (EDC), which is computed by reversely integrating the squared RIR,

$$EDC(t) = G\int_t^{\infty} h^2(\tau)d\tau \quad (4)$$

where G is a constant related to excitation level. Then, RT can be obtained by estimating of EDC's decay rate over $[-5dB, -35dB]$ and computing the time taken to decay by 60 dB. ISO 3382 specifies the preceding measurement method as a standard. However, this method is not applicable in our case. Firstly, the RIRs we collect are very noisy. The noise stems from both the measurement equipment and the background. Noise dominates and stretches the energy decay curve, as illustrated in Fig. 5. Secondly, the positions of the speaker and the microphone are very close to each other on laptops or smartphones, which results in a very strong direct feed-through. It appears as a steep drop at the beginning of both RIR and EDC. However, this segment makes no contribution to the calculation of RT, since the direct sound energy depends only on the distance between speaker and microphone and is independent of rooms' properties. The only EDC segment useful for RT calculation is where reverberation dominates, but it is very short as illustrated in Fig. 5. The sound energy decays by less than 10 dB before overwhelmed by noise. Therefore, a new noise compensation method is needed to extract RT from very noisy RIRs.

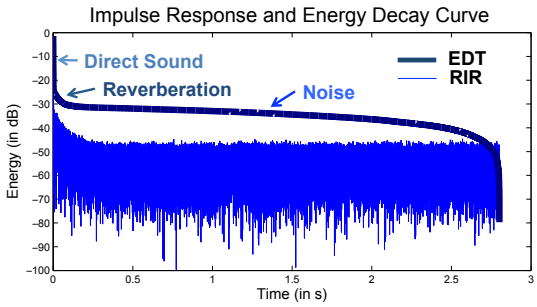

Fig. 5. Impulse response and energy decay curve. EDC segments corresponding to direct sound domination, reverberation domination, noise domination are annotated.

We propose the Noise Adaptive Extraction of Reverberation (NAER) algorithm to estimate RT from noisy RIRs. NAER first estimates the noise level of the environment and then defines RT as the time taken for sound energy decays to noise level (Fig. 6); therefore, RT is well-defined regardless of noise level. A pseudocode of NAER is provided in Algorithm 1. We calculate RT of the rooms listed in Table. I using NEAR, where the corresponding parameters *PerNoise*, *BondP* and Th are set to be 90%, the midpoint of the RIR and $0.5dB$. The result is consistent with volumes and absorption properties of rooms, as illustrated in Fig. 7.

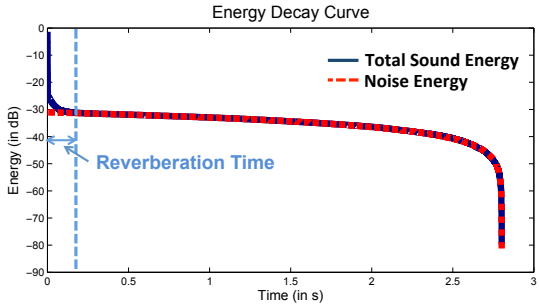

Fig. 6. NAER estimate the noise energy and defines RT as the time taken for total sound energy decay to noise energy.

Algorithm 1 Pseudo-code of NAER Algorithm

function NAER(*RIR*, *PerNoise*, *BondP*, Th)
  Inputs:
    RIR: room impulse response of length L
    PerNoise: the last *PerNoise* portion of RIR used to estimate noise level
    BondP: the bonding point defined where sound energy meets noise
    Th: threshold to define reverberation time
  Output:
    RT: reverberation time
  Noise Estimation:
    $NoiseLevel \leftarrow RIR(PerNoise : end)$
  Pseudo Noise Energy Curve Calculation:
    $PseudoNoiseEnergy \leftarrow$ inverse integrate $NoiseLevel$
    $SoundEnergy \leftarrow$ inverse integrate RIR
    $PseudoNoiseEnergy \leftarrow PseudoNoiseEnergy + SoundEnergy(BondP) - PseudoNoiseEnergy(BondP)$
  Reverberation Time Extraction:
    for $FindInd \in \{1, \cdots, L\}$ **do**
      if $SoundEnergy(FindInd) - PseudoNoiseEnergy(FindInd) < Th$ **then**
        $RT \leftarrow FindInd$; **break**
      end if
    end for
end function

We also extract RT in different octave bands in order to take into account the frequency-dependent absorption properties of the room. Each octave band is identified via its center frequency, e.g., RT 500Hz represents the RT extracted from the frequency band $[\frac{500}{\sqrt{2}}\text{Hz}, 500\sqrt{2}\text{Hz}]$. In addition, early to total sound energy ratio (D50), early to late arriving sound energy ratio (C50) and center time of the squared impulse response (TS) are also used as energetic features. Generally speaking, these parameters describe where the sound energy is concentrated along the timeline. The dominance of early energy is an indicator for a smaller volume or a low sound absorption (Fig. 8). D50, C50, TS can be computed by the following formulas: $D_{50} = \frac{\int_0^{0.05s} h^2(t)dt}{\int_0^{\infty} h^2(t)dt}$, $C_{50} = 10log\left(\frac{D_{50}}{1-D_{50}}\right)$, $TS = \frac{\int_0^{\infty} th^2(t)dt}{\int_0^{\infty} h^2(t)dt}$.

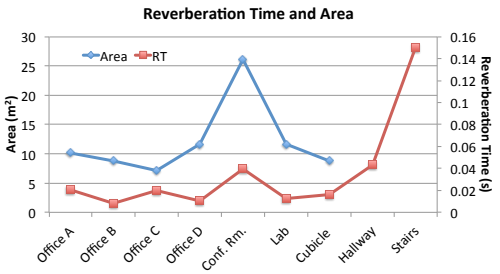

Fig. 7. RT given by NAER varies in the same trend as the area values. Particularly, we do not have access to precise area values of hallway and stairs, but they are much larger than other locations investigated and these two locations also have much larger RTs as is shown here.

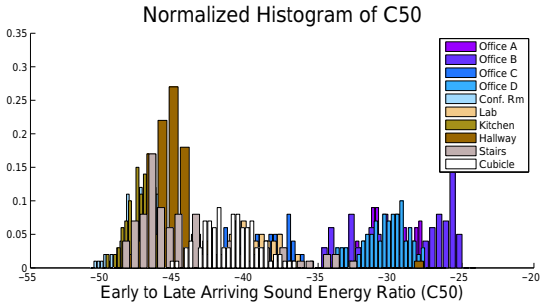

Fig. 8. Normalized histogram of early to late arriving sound energy ratio. Sound energy in rooms with smaller volumes tends to be dominated by early energy, while C50 for closed spaces with larger volumes or open spaces tends to be smaller.

VI. Performance Evaluation

A. Distinctiveness

In order to support localization, fingerprints should have a good separation between distinctive areas. In other words, there should be a one-to-one mapping from a room label to a feature distribution. The dissimilarity of distributions can be measured by Jensen-Shannon (JS) divergence, which is known for its capability of measuring a divergence between more than two probability distributions. JS divergence is defined as

$$JS(P_1, \cdots, P_M) = \sum_{i=1}^{M} \pi_i KL(P_i || \bar{P}) \tag{5}$$

where $\bar{P} = \sum_{i=1}^{M} \pi_i P_i$ is the mixed distribution, π_i represents the weight for the distribution P_i, $\pi_i \in [0,1]$, $\sum_{i=1}^{M} \pi_i = 1$.

$KL(P_i||\bar{P})$ is the Kullback-Leibler divergence, defined as $KL(P_i||\bar{P}) = \sum_x \log\left(\frac{P_i(x)}{\bar{P}(x)}\right) P_i(x)$. JS divergence is a weighted sum of KL divergence and measures the distinctiveness of multiple distributions by considering how far each of the distributions deviates from the mixed distribution. The larger the JS divergence is, the better the separability the feature has achieved.

We verify the distinctiveness of a certain feature by classical permutation test. The idea is to randomly permute the labels of room labels and each time obtain a JS divergence. The null hypothesis is that the observed JS divergence for a given feature is independent of the room labeling, namely

$$H_0^{Feat.} : JS_{Observed}^{Feat.} = JS_{Permuted}^{Feat.} \quad (6)$$

where $H_0^{Feat.}$ denotes the null hypothesis for a certain feature. If the observed JS divergence significantly deviates from the mean of the JS divergence distribution in permutation test, we can reject the null hypothesis, i.e., the feature is distinctive for different locations. The result of the permutation test on all the features in our feature pool is presented in Fig. 9. The error bar specifies the quadruple standard deviation below and above the mean of JS divergence distribution in permutation test. As can be seen, the observed JS divergence is significantly larger than the distribution obtained in the permutation test. The result of p-value of significance testing is 0.0002, which shows that the probability of obtaining a JS divergence as extreme as observed under the null hypothesis is extremely small; therefore, the null hypothesis is rejected. We conclude that the features presented above can achieve high separability for different locations.

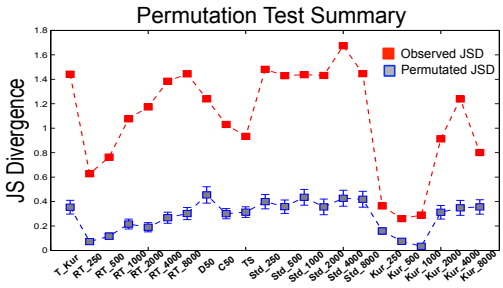

Fig. 9. Permutation test summary. The observed JS divergence is significantly larger than the distribution of JS divergence obtained in permutation test and the null hypothesis should be rejected.

Within the context of localization, we further evaluate the feature distinctiveness by measuring the localization accuracy. The Sequential Floating Forward Selection (SFFS) algorithm was used to select a set of features to minimize the prediction error [**?**]. One essential reason for feature selection is to avoid overly complex models with respect to the number of features employed, since a sparse model is more robust to changes under different circumstances. For instance, we expect slight variations for feature distribution at different background noise levels, or at different times, as we will discuss in the next two sections. A robust set of features should be able to still distinguish the location from other in such conditions.

We use classification error as the criterion for SFFS and the Naive Bayes Classifier (NBC) is implemented here for classification. Fig. 10 shows the classification confusion matrix for the room identification using optimal feature set listed in Table II. The overall accuracy is 97.8%. When all features are used for classification, the overall accuracy is 95.9%.

	Office A	Office B	Office C	Office D	Conf. Rm	Lab	Kitchen	Hallway	Stairs	Cubicle
Office A	98			2			1			
Office B	1	97		4						
Office C			100							
Office D		3		93				1		
Conf. Rm					100					
Lab	1			1		99	2		2	
Kitchen							94			
Hallway								99		
Stairs						1	3		98	
Cubicle										100

Fig. 10. Confusion matrix for Experiment A. The features used are from optimal set list in Table 3. The overall accuracy is 97.8%.

TABLE II
OPTIMAL FEATURE SET CREATED BY SFFS

Temporal Feat.	Kur.
Spectral Feat.	Std. 1000 Hz, Std. 2000 Hz, Std. 4000 Hz, Std. 8000 Hz, Kur. 1000 Hz, Kur. 2000Hz, Kur. 4000 Hz, Kur. 8000 Hz
Energetic Feat.	RT 1000 Hz, RT 2000 Hz, RT 4000 Hz, RT 8000 Hz, C50, D50, TS

B. *Noise-robustness*

Generally speaking, any acoustic localization system exploits the location information hidden behind the recordings. Inevitably, there exist some transient noise that are independent of the position and bring ambiguity to location estimation. For instance, speech noise leads to an over 20% accuracy drop in the localization method based on ambient background sound sensing [9]. Therefore, noise-robustness is a challenging issue in acoustic localization system. Our localization technique leverages information from RIRs. They are computed using MLS excitation and cross-correlation technique. Since the phase spectrum of MLS is strongly erratic with a uniform density of probability in the $[-\pi, +\pi]$, transient noise like clicks, footsteps, etc. will be randomized and transformed into benign noise distributed evenly over the entire impulse response. Therefore, MLS-based RIR measurement should be expected to be immune to extraneous noise of all kinds theoretically. We design *Experiment B* to test the noise-robustness of SoundLoc, where test and training set are noisy and quiet samples collected in the conference room, respectively. The accuracy of labeling the conference room is used as the indicator of noise-robustness. The result is presented in Table III. The

accuracy is poor when all features or optimal features determined from Experiment A are used for classification. This result shows that in practice the transient noise cannot be fully weakened by using MLS technique and some features in our feature pool are sensitive to noise. Again, we use SFFS to determine which feature is least sensitive to noise. The reselected features are listed in Table IV. They tend to exclude voice band, which is approximately 80-260 Hz. When using these selected features for classification, 98% accuracy can be achieved for labeling the conference room.

TABLE III
RESULTS SUMMARY OF *Experiment B*

Location	Accuracy		
	All Feat.	Exp. A Feat.	Reselected Feat.
Conf. Rm.	15%	24%	98%

TABLE IV
NOISE-ROBUST FEATURE SET CREATED BY SFFS

Spectral Feat.	Std. 500 Hz, Std. 1000 Hz, Std. 2000 Hz, Std. 4000 Hz, Kur. 1000 Hz, Kur. 2000Hz
Energetic Feat.	RT 500 Hz, RT 1000 Hz, RT 8000Hz, C50, D50, TS

C. Time-invariance

A useful fingerprint should be relatively stationary over time. We design *Experiment C* to test the time-invariance of SoundLoc, where test and training sets are from separate visits. The result is summarized in Table V. In general, using data from completely different visits for training results in a slightly lower accuracy than that when training and testing data come from the same visit. For Office B, the accuracy suffers from a dramatic fall when different visit samples are used for training. However, for stairs and cubicle, the accuracy remains above 99%. The reason for this is that stairs and cubicle are very different from other locations investigated in our paper, while the 4 offices in our experiment have similar geometry, wall materials and furnishings. The variation of features leads to confusion among very similar locations. We also test the features that are chosen by SFFS in *Experiment A*. The accuracy using this feature set is lower than that when all features are used for classification. That's because some features that are invariant in longer time scale but do not lead to the best accuracy are excluded during feature selection in *Experiment A*. We reselect the most time-stationary features using SFFS, as listed in Table VI. Higher than 93% accuracy can be achieved with the reselected features.

D. Comprehensive Feature Selection

The previous three experiments consider three different data collection and testing scenarios. *Experiment A* is when the distribution of testing and training data do not differ by noise or time. *Experiment B* and *C* consider the effect of noise and time transition on RIRs respectively. We also

TABLE V
RESULTS SUMMARY OF *Experiment C*

Location	Training Type	Accuracy		
		All Feat.	Exp. A Feat.	Reselected Feat.
Office B	Different Visit	79%	76%	95%
	Same Visit	93%	97%	93%
Stairs	Different Visit	99%	98%	99%
	Same Visit	99%	98%	99%
Cubicle	Different Visit	100%	100%	100%
	Same Visit	99%	100%	99%

TABLE VI
TIME-INVARIANT FEATURE SET CREATED BY SFFS

Temporal Feat.	Kur.
Spectral Feat.	Std. 250 Hz, Std. 500 Hz, Std. 1000 Hz, Std. 2000 Hz, Std. 4000 Hz, Std. 8000 Hz, Kur. 250 Hz, Kur 500 Hz, Kur. 1000 Hz, Kur. 2000 Hz
Energetic Feat.	RT 250 Hz, RT 500Hz, RT 2000 Hz, RT 4000 Hz, RT 8000 Hz, C50, D50, TS

conducted an additional experiment that considers all the settings by combining the noise- and time-corrupted data and use the mixed data for feature selection. In reality it is likely that the data is collected and tested under any one of the four scenarios.

Table VII organizes the features by the number of time they are selected in each case. Features that are consistently selected, such as RT 8000 Hz, TS, and Std. 4000 Hz, are very likely to perform well in practice, since they stand the test of noise-robustness and time-invariance. Those who are selected more than two times are also good features since their inclusion can enhance the classification performance in most situations. We notice that some low frequency features are not likely to be selected, since they are often susceptible to talking and ambient noises. As a guideline, the user is advised to include features sequentially from the first row to the last row in Table VII, depending on the model complexity. Usually, sparse model is good for generality and more complex model is good if the situation is not highly dynamic.

TABLE VII
COMPREHENSIVE FEATURE SELECTION TABLE

# time selected	Features
4 out of 4	RT 8000 Hz, TS, Std. 4000 Hz, Kur. 1000 Hz, Kur. 2000 Hz
3 out of 4	Time Kur., RT 2000 Hz, RT 4000 Hz, C50, D50, Std. 500 Hz, Std. 1000 Hz, Std 2000 Hz, Std 8000 Hz, Kur. 1000 Hz, Kur. 8000 Hz
2 out of 4	RT 250 Hz, RT 500 Hz, RT 1000 Hz, Kur 4000 Hz, Kur 8000 Hz

E. Energy Footprint Optimization

In this section, we consider the overall energy footprint for modeling a specific location, i.e., the number of samples

that are required for reliably labeling this place. There are several reasons to optimize the number of training sets for SoundLoc. Firstly, training labels are often costly and time-consuming to obtain. Secondly, it requires storage space on the mobile platform, so a large set of training samples might limit the number of places in the memory. Also, more training samples demand more computational power, which might represent a bottleneck on the battery-powered device. To study the effects of training size on classification accuracy, we vary the size of the training sets to train an array of popular classifiers and plot the results in Fig. 11.

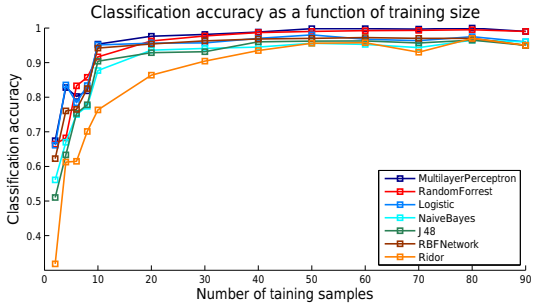

Fig. 11. The effect of the number of training samples on the classification accuracy for various algorithms. All the models are implemented by the Weka machine learning toolkit [**?**]. The required number of training samples are selected randomly from the training sets. The rest of the samples are used as the testing set. All the features in our feature pool are used for classification.

As it can be seen, the classification accuracy generally improves as the number of training samples increases. Most methods converge to an optimal classification rates when the number of training samples is from ten (10) to twenty (20). For the top algorithms such as Multilayer Perceptron and Random Forrest, the accuracy achieves 95.33% and 91.67% respective with only ten (10) samples. The satisfactory performance with only limited number of training samples is attributed to the separability, noise robustness, and time invariance of the sound features. Relaxing the requirement of training samples directly benefits the energy efficiency of SoundLoc, as well as convenience to implement.

VII. Conlustion

We have presented SoundLoc, a room identification system exploiting the acoustic properties of the room. The acoustic properties are described quantitatively by various features extracted from RIR. We build a cheap MLS-based RIR measurement system using internal speakers and microphones on laptops. The NAER algorithm is developed to extract features from the noisy RIRs. The algorithm is shown to be effective to extract RT when the sound energy decay is dominated by direct sound and noise. Using this measurement system, we collect more than 1000 RIR samples in different locations, with different noise background and time stamps. The acoustic features we extracted are shown to be distinctive, robust and efficient to compute. The overall accuracy of 97.8% has been achieved for 10 rooms' identification. Moreover, the training sample size can be reduced to 10 samples while 95.3% accuracy can still be achieved.

References

[1] V. L. Erickson, S. Achleitner, and A. E. Cerpa, "Poem: Power-efficient occupancy-based energy management system," in *Information Processing in Sensor Networks (IPSN), 2013 ACM/IEEE International Conference on*. IEEE, 2013, pp. 203–216.

[2] Z.-N. Zhen, Q.-S. Jia, C. Song, and X. Guan, "An indoor localization algorithm for lighting control using rfid," in *Energy 2030 Conference, 2008. ENERGY 2008. IEEE*. IEEE, 2008, pp. 1–6.

[3] A. Williams, D. Ganesan, and A. Hanson, "Aging in place: fall detection and localization in a distributed smart camera network," in *Proceedings of the 15th international conference on Multimedia*. ACM, 2007, pp. 892–901.

[4] R. Want, A. Hopper, V. Falcao, and J. Gibbons, "The active badge location system," *ACM Transactions on Information Systems (TOIS)*, vol. 10, no. 1, pp. 91–102, 1992.

[5] A. Haeberlen, E. Flannery, A. M. Ladd, A. Rudys, D. S. Wallach, and L. E. Kavraki, "Practical robust localization over large-scale 802.11 wireless networks," in *Proceedings of the 10th annual international conference on Mobile computing and networking*. ACM, 2004, pp. 70–84.

[6] 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*, vol. 2. Ieee, 2000, pp. 775–784.

[7] Y. Jiang, X. Pan, K. Li, Q. Lv, R. P. Dick, M. Hannigan, and L. Shang, "Ariel: Automatic wi-fi based room fingerprinting for indoor localization," in *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*. ACM, 2012, pp. 441–450.

[8] M. Azizyan, I. Constandache, and R. Roy Choudhury, "Surroundsense: mobile phone localization via ambience fingerprinting," in *Proceedings of the 15th annual international conference on Mobile computing and networking*. ACM, 2009, pp. 261–272.

[9] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, "Indoor localization without infrastructure using the acoustic background spectrum," in *Proceedings of the 9th international conference on Mobile systems, applications, and services*. ACM, 2011, pp. 155–168.

[10] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, "Soundsense: scalable sound sensing for people-centric applications on mobile phones," in *Proceedings of the 7th international conference on Mobile systems, applications, and services*. ACM, 2009, pp. 165–178.

[11] N. Peters, H. Lei, and G. Friedland, "Name that room: room identification using acoustic features in a recording," in *Proceedings of the 20th ACM international conference on Multimedia*. ACM, 2012, pp. 841–844.

[12] N. R. Shabtai, Y. Zigel, and B. Rafaely, "Estimating the room volume from room impulse response via hypothesis verification approach," in *Statistical Signal Processing, 2009. SSP'09. IEEE/SP 15th Workshop on*. IEEE, 2009, pp. 717–720.

[13] G.-B. Stan, J.-J. Embrechts, and D. Archambeau, "Comparison of different impulse response measurement techniques," *Journal of the Audio Engineering Society*, vol. 50, no. 4, pp. 249–262, 2002.

[14] D. D. Rife and J. Vanderkooy, "Transfer-function measurement with maximum-length sequences," *Journal of the Audio Engineering Society*, vol. 37, no. 6, pp. 419–444, 1989.

[15] J. Borish and J. B. Angell, "An efficient algorithm for measuring the impulse response using pseudorandom noise," *Journal of the Audio Engineering Society*, vol. 31, no. 7/8, pp. 478–488, 1983.

[16] J. J. Jetzt, "Critical distance measurement of rooms from the sound energy spectral response," *The Journal of the Acoustical Society of America*, vol. 65, 1979.

[17] A. H. Moore, M. Brookes, and P. A. Naylor, "Roomprints for forensic audio applications," in *WASPAA*. IEEE, 2013.

[18] H. Kuttruff, *Room Acoustics*. London: Taylor and Francis, 2000.

[19] P. Somol, P. Pudil, J. Novoviov, and P. Paclk, "Adaptive floating search methods in feature selection. pattern recognition letters," *Pattern recognition letters*, vol. 20, 1999.

[20] M. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, and I. H. Witten, "The weka data mining software: An update," 2009.