EchoTag: Accurate Infrastructure-Free Indoor Location Tagging with Smartphones

Yu-Chih Tung and Kang G. Shin
The University of Michigan
Email: {yctung,kgshin}@umich.edu

ABSTRACT

We propose a novel mobile system, called EchoTag, that enables phones to tag and remember indoor locations without requiring any additional sensors or pre-installed infrastructure. The main idea behind EchoTag is to *actively* generate acoustic signatures by transmitting a sound signal with a phone's speakers and sensing its reflections with the phone's microphones. This *active* sensing provides finer-grained control of the collected signatures than the widely-used passive sensing. For example, because the sensing signal is controlled by EchoTag, it can be intentionally chosen to enrich the sensed signatures and remove noises from useless reflections. Extensive experiments show that EchoTag distinguishes 11 tags at 1cm resolution with 98% accuracy and maintains 90% accuracy even a week after its training. With this accurate location tagging, one can realize many interesting applications, such as automatically turning on the silent mode of a phone when it is placed at a pre-defined location/area near the bed or streaming favorite songs to speakers if it is placed near a home entertainment system. Most participants of our usability study agree on the usefulness of EchoTag's potential applications and the adequacy of its sensing accuracy for supporting these applications.

Categories and Subject Descriptors

H.3.4 [**Information Storage and Retrieval**]: Systems and Software

General Terms

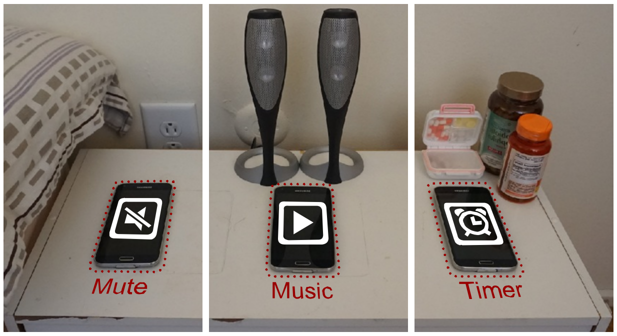
Design, Measurement, Performance, Algorithms

Keywords

Localization, Sound, Fingerprinting, Mobile Phones

1. INTRODUCTION

Imagine one day, the silent mode of a phone is automatically activated in order to avoid disturbing a user's sleep when the phone is placed near the bed. Likewise, favorite songs are streamed to speakers whenever the phone is placed near a stereo or a predefined timer/reminder is set if the phone is near a medicine cabinet. This kind of applications is known as context-aware computing or indoor geofencing which provides a natural combination of function and physical location. However, such a function–location combination is still not pervasive because smartphones are not yet able to sense locations accurately enough without assistance of additional sensors or pre-installed infrastructure.



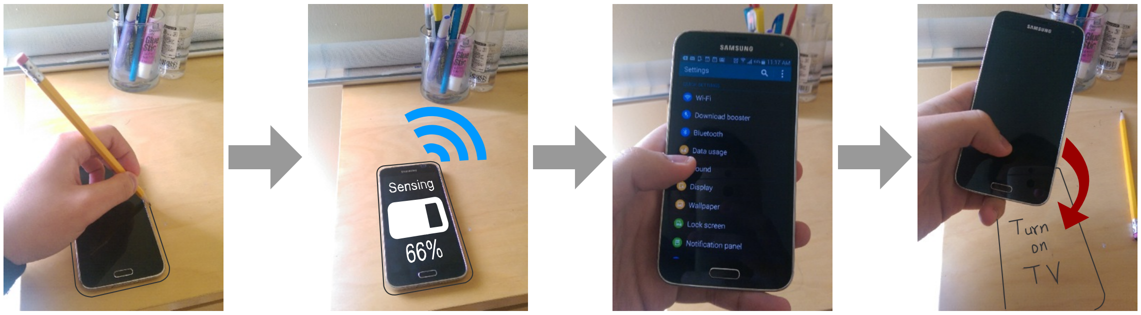
(a) Auto silent mode (b) Auto music (c) Auto timer

Figure 1—Candidate applications of EchoTag. Silent mode is automatically activated when the phone is placed on a drawn box, named *(echo) tag*, near the bed. Favorite songs are streamed to speakers or a predefined timer is automatically set when the phone is placed at other nearby tags.

Existing localization systems are unable to provide this type of functionality for two reasons. First, they usually rely on *passively* recorded WiFi, FM, or background acoustic signals, and can only achieve about room- or meter-level accuracy. Nevertheless, the above-mentioned applications need more accurate location sensing, e.g., both streaming music and setting silent mode might take place in the same room or even on the same table as shown in Fig. 1. Second, more accurate (i.e., with error of a few cm) location sensing with light recording or acoustic beacons requires a pre-installed infrastructure. The cost of such an infrastructure and the ensuing laborious calibrations make its realization expensive or difficult, especially for personal use.

In this paper, we propose a novel location tagging system, called EchoTag, which enables phones to tag and remember indoor locations with finer than 1cm resolution and without requiring any additional sensors or pre-installed infrastructure. The main idea behind EchoTag is to *actively* render acoustic signatures by using phone speakers to transmit sound and phone microphones to sense its reflections. This *active* sensing provides finer-grained control of the collected signatures than the commonly-used passive sensing. For example, EchoTag emits sound signals with different delays at the right channel to enrich the feature space for sensing nearby locations, and exploits the synchronization between the sender and the receiver as an anchor to remove interferences/reflections from objects outside the target area/locations. Moreover, this active sensing relies only on built-in sensors available in commodity phones, thus facilitating its deployment. Note that EchoTag is not designed to replace any localization system since it can only remember the locations where it had been placed before, rather than identifying arbitrary indoor locations. However, this fine-grained location sensing for remembering location tags can enable many important/useful applications that have not yet been feasible due to large location sensing errors or the absence of pre-installed infrastructure.

MobiCom'15, September 7–11, 2015, Paris, France.
© 2015 ACM. ISBN 978-1-4503-3619-2/15/09…$15.00.
DOI: http://dx.doi.org/10.1145/2789168.2790102.



(a) Draw (echo) tags (b) Sense acoustic signatures (c) Select mapped applications (d) Replay recorded tags

Figure 2—Four steps of using `EchoTag`. The user first draws the contour of target locations/areas with a pencil, then commands the phone to sense the environment. After sensing the environment, a combination of applications and functions to be performed at this location is selected. Finally, the user automatically activates the selected applications/functions by simply placing his phone back within the contoured area. The contoured areas are thus called *(echo) tags*.

Fig. 2 demonstrates a 4-step process to set up and use EchoTag. The first step is to place the phone at the target location and draw the contour of the phone with a pencil. This contour is used as a marker for users to remember the target location, which is called an *(echo) tag*. (Tags can also be drawn on papers pasted on target locations.) Then, EchoTag generates and records the surrounding signatures of this target location. The next step is to select the applications/functions being combined and associated with this tagged location. Finally, users can easily activate the combined applications/functions by placing their phones back in the drawn tags. In summary, EchoTag embeds an invisible trigger at physical locations that the phone remembers what to do automatically.

We have implemented EchoTag as a background service in Android and evaluated the performance by using Galaxy S5 and other mobile devices. Our experimental evaluation shows that commodity phones equipped with EchoTag distinguish 11 tags with 98% accuracy even when tags are only 1cm apart from each other and achieves 90% accuracy based on the trace collected a week ago. Since EchoTag only utilizes existing sensors in commodity phones, it can be easily implemented and deployed in other mobile platforms. For example, we also implemented and evaluated EchoTag on iOS (with traces classified in Matlab) but omitted the results due to space limitation. Our usability study of 32 participants also shows that more than 90% of participants think the sensing accuracy and prediction delay of EchoTag are useful in real life. Moreover, about 70% of the participants agree that the potential applications in Fig. 1 can save time and provide convenience in finding and activating expected functions.

This paper makes the following four contributions:

- The first indoor location tagging achieving 1cm resolution by using commodity phones;
- Demonstration of the ability of active sensing to enrich the feature space and remove interferences;
- Implementation of EchoTag in Android without any additional sensors and/or pre-installed infrastructure; and
- Evaluation of EchoTag, showing more than 90% accuracy even a week after locations were tagged.

The remainder of this paper is organized as follows. Section 2 discusses the related work in indoor location sensing. Section 3 provides an overview of EchoTag. Sections 4–6 describe the design of acoustic signature and classifiers. The implementation details are provided in Section 7, and the performance of EchoTag and its real-world usability are evaluated in Sections 8 and 9, respectively. We discuss future directions in Section 10 and conclude the paper in Section 11.

System	Resolution	Infrastructure	Signature
SurroundSense [3]	room-level	No	Fusion
Batphone [25]	room-level	No	Sound
RoomSense [23]	300cm	No	Sound
Radar [4]	400cm	Existing	WiFi
Horus [30]	200cm	Existing	WiFi
Geo [8]	100cm	No	Geomagnetism
FM [7]	30cm	Existing	FM
Luxapose [15]	10cm	Additional	Light
Cricket [22]	10cm	Additional	Sound/WiFi
Guoguo [17]	6–25cm	Additional	Sound
`EchoTag`	**1cm**	**No**	**Sound**

Table 1—Existing indoor location sensing systems.

2. Related Work

Indoor localization is a plausible entry to location tagging. The existing localization systems are summarized in Table 1. The most popular methods used for indoor localization, such as Radar [4] and Horus [30], sense locations based on WiFi-signal degradation. Their main attractiveness is the reliance on widely-deployed WiFi, hence requiring a minimal deployment effort. However, severe multipath fading of WiFi signals makes WiFi-signature-based localization achieve only room-level accuracy. To overcome the instability of WiFi signatures and increase the accuracy of indoor localization, researchers have also explored other sources of signatures. For example, the authors of [8] adopted the readings of geo-magnetism which varies with location due to the disturbance of steel structure of buildings. FM radio is also adopted to increase the sensing accuracy of WiFi-based localization [7]. Batphone [25] determines the room locations by sensing the background acoustic noise. Unfortunately, even with these improvements, localization systems relying on passive sensing of the environment can only achieve meter-level resolution. Moreover, passively sensing the environment suffers greatly when the environment changes. For example, as shown in [25], the signature of background acoustic noise changes dramatically when the climate control (HVAC) system was shut off for maintenance, and WiFi RSSI is known to change significantly when the transmit power control in an Access Point is enabled [7].

As shown in Table 1, a few systems, such as Luxapose [15], Cricket [22] and Guoguo [17], provide indoor localization with a few cm resolution, thus enabling accurate location tagging. However, these fine-grained localization systems can only be realized with the help from a pre-installed infrastructure. For example, Luxapose requires replacement of ceiling lamps by programmed LED and Guoguo & Cricket require customized WiFi/acoustic beacons around the building. Even if the cost of each additional sensor might be affordable, the aggregated cost and the laborious calibration required for this deployment are still too high to be attractive/feasible for real-world deployment. Even with the pre-installed infrastructure, the localization error is still around 10cm. The localization error can be reduced further by using antenna or microphone arrays [10, 29], but these advanced sensors are not available in commodity phones. In contrast, EchoTag achieves location sensing with 1cm resolution without any pre-installed infrastructure or additional advanced sensors. Location tagging can also be realized by deploying NFC tags [2], but EchoTag makes this functionality realizable in all commodity phones — such as HTC butterfly, the latest Xiaomi 4, and iOS phones[1] — that are commonly equipped with microphones and speakers, but not NFC chips. Note that EchoTag is *not* designed to replace any localization system since it can only remember the locations where it had been placed before, rather than identifying arbitrary indoor locations like [15, 17, 22]. However, as the results shown in this paper and the feedbacks from the participants of our usability study, this fine-grained location sensing for remembering tags can be used to enable many important/useful applications that have not been feasible before, due to large location sensing errors or lack of installed infrastructure.

EchoTag senses locations based on acoustic signatures. Acoustic signals have been studied widely since they are readily available in commodity phones. For example, SurroundSense [3] and Auditeur [18] classify user behaviors based on background noise. Skinput [12], TapSense [11] and SufaceLink [9] provide new commuter–human interactions based on acoustic reflections from human skin, fingertip, or contacted surface. UbiK [28] provides a input method with acoustic signatures in response to touching different positions on a table. All of these use similar acoustic signatures (e.g., resonances) as in EchoTag, but EchoTag is the first to use it for accurate indoor location tagging. Additional novelties of EchoTag include its *active* generation of acoustic signatures and enrichment of signatures by emitting sound signals with different delays.

The closest to EchoTag are Touch & Active [19], Symbolic Object Localization [14], and RoomSense [23], all of which also actively generate acoustic signals and record their signatures but for different purposes. Touch & Active [19] uses the same multipath signature to identify how the user touches an object equipped with piezo speakers and microphones. Commodity phones were mentioned as a potential interface for Touch & Active, but no evaluation was provided. The authors of [14] use sound absorption by the touched surface as a feature to identify symbolic locations of a phone — i.e., in a pocket, on a wood surface or a sofa — which is unable to detect nearby locations on the same surface. RoomSense [23] also uses the sound reflections from environments to identify different rooms. However, since only the compressed analytical feature (e.g., Mel Frequency Cepstral Coefficient) is used, its sensing resolution is larger than 9m^2, thus becoming unable to distinguish nearby tags. In this paper, we implement a novel accurate location tagging system based on actively generated acoustic signatures via built-in phone sensors.

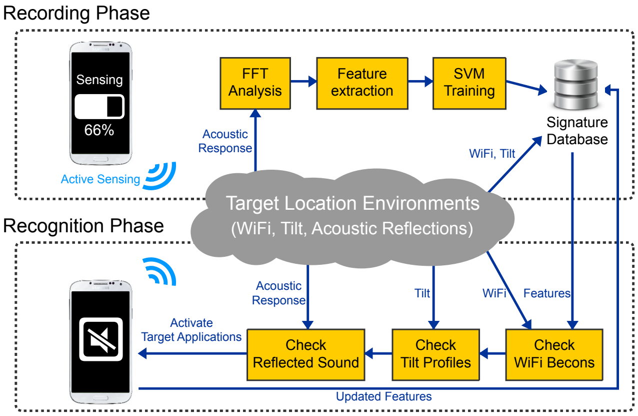


Figure 3—System overview. Locations are sensed based on acoustic reflections while the tilt/WiFi readings are used to determine the time to trigger acoustic sensing, thus reducing the energy consumption of the sensing process.

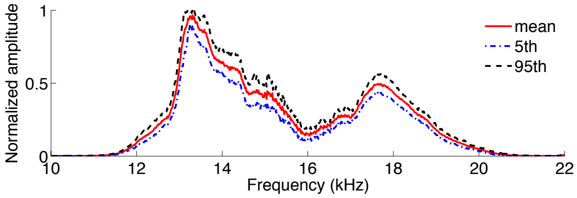


Figure 4—An example of acoustic signatures. The received attenuation of a flat frequency sweep is uneven over different frequencies. The result is an average of 100 trials over 1 minute.

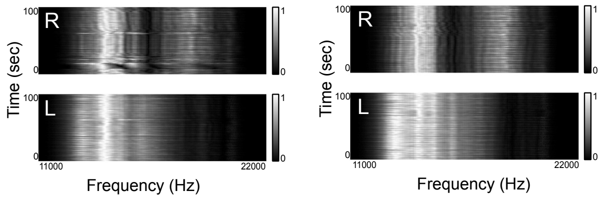
3. System Overview

Fig. 3 gives an overview of EchoTag which is composed of recording and recognition phases. In the recording phase, multiple short sequences of sound signals will be emitted from the phone speakers. Each sequence is repeated a few times with different delays between left and right channels to enrich the received signatures as we will discuss in the following sections. The reading of built-in inertial sensors is also recorded for further optimization. After recording the signature, the selected target application/function and the collected signatures are processed and saved in the device's storage. In the recognition phase, the phone will continuously check if the WiFi SSID and the tilt of the phone match the collected signatures. If the tilt and WiFi readings are similar to one of the recorded target locations, then the same acoustic sensing process is executed again to collect signatures. This new collected signature is compared with the previous records in the database using a support vector machine (SVM). If the results match, the target application/function will be automatically activated.

4. Acoustic Signature

EchoTag differentiates locations based on their acoustic signatures, characterized by uneven attenuations occurring at different frequencies as shown in Fig. 4. Note that EchoTag does not examine the uneven attenuations in the background noise but those in the sound emitted from the phone itself. For example, as shown in Fig. 4, the recorded responses of a frequency sweep from 11kHz to 22kHz are not flat but have several significant degradations at certain frequencies. The characteristics of this signature at different locations can be observed in Fig. 5 where the phone is moved 10cm away from its original location. In what follows, we will un-

[1] Apple locks iPhone 6 NFC to Apple Pay:
http://www.cnet.com/news/apple-locks-down-iphone-6-nfc-to-apple-pay/



(a) Location A (b) Location B

Figure 5—Frequency responses at nearby locations. Responses varies with location (i.e., the distribution of light and dark vertical lines) and this is used as a feature for accurate location tagging.

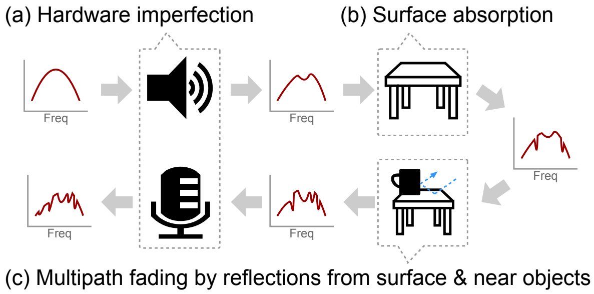

(a) Hardware imperfection
(b) Surface absorption
Freq
Freq
Freq
Freq
Freq
(c) Multipath fading by reflections from surface & near objects

Figure 6—Causes of uneven attenuation. During the recording of emitted sound, hardware imperfection of microphones/speakers, absorption of touched surface materials and multipath reflections from nearby objects incur different degradations at different frequencies. Only the degradation caused by multipath reflections is a valid signature for sensing locations even in the same surface.

earth the causes of this phenomenon and describe how to exploit this feature for EchoTag's accurate location tagging.

4.1 Causes of Uneven Attenuation

There are three main causes of this uneven attenuation: (a) hardware imperfection, (b) surface's absorption of signal, and (c) multipath fading caused by reflection. As shown in Fig. 6, when sound is emitted from speakers, hardware imperfections make the signal louder at some frequencies and weaker at other frequencies. These imperfections have been identified and used as a signature to track people's smartphones for the purpose of censorship [31]. After the emitted sound reaches the surface touched by the phone, the surface material absorbs the signal at some frequencies. Different materials have different absorption properties, thus differentiating the surface on which the phone is placed [14]. Then, when the sound is reflected by the touched surface and the surrounding objects, the combination of multiple reflections make received signals constructive at some frequencies while destructive at other frequencies. This phenomenon is akin to multipath (frequency-selective) fading in wireless transmissions. For example, if the reflection of an object arrives at microphones t milliseconds later than the reflection from the touched surface, then the signal component at $10^3/2t$ Hz frequency of both reflections will have opposite phases, thus weakening their combined signal. This multipath property of sound has been shown and utilized as a way to implement a ubiquitous keyboard interface [28]. When reflections reach the phone's microphone, they will also degrade due to imperfect microphone hardware design.

For the purpose of accurate location tagging, EchoTag relies on the multipath fading of sound among the properties mentioned above as this is the only valid signature that varies with location even on the same surface. In what follows, we will introduce the challenges of extracting this feature and describe how EchoTag meets them.

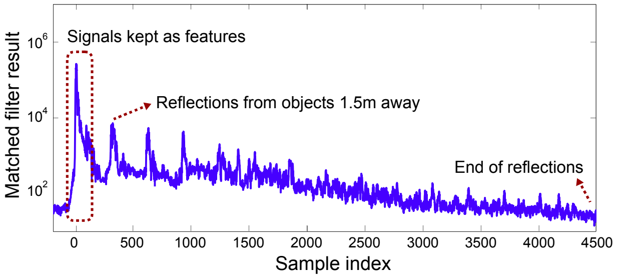


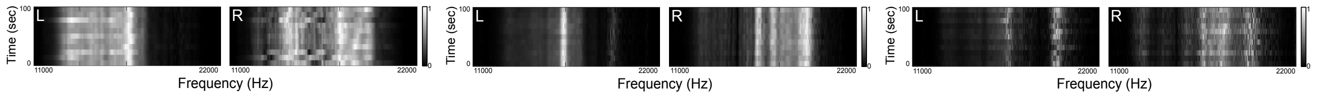
Figure 7—Characteristics of reflections. A matched filter is used to identify the reflections of a 100-sample chirp. Only first 200 samples after the largest peak are kept as a feature in EchoTag, excluding reflections from objects farther than 86cm away.

4.2 Sound Selection

The selection of parameters for the emitted sound is critical for EchoTag to extract valid multipath signatures. According to the guideline of Android platforms,[2] 44.1kHz is the most widely supported sample rate for Android phones, so the highest frequency that can be sensed is about 22kHz. Studies have shown that humans can hear signals of frequency up to 20kHz [21]. It is thus desirable to make the emitted sound inaudible (to avoid annoyance) by sensing 20 to 22kHz. But from our preliminary experiments on commodity phones, we found that the signal responses in this high-frequency range are not strong enough to support accurate indoor location tagging due to the imperfect hardware design that causes significant degradation of signal strength in this high-frequency range. Based on the experiments in [16], certain phones' microphones receive signals with 30dB less strength at 22kHz. This phenomenon is even worse if the degradation of speakers is accounted for. Thus, we choose the chirp (i.e., frequency sweep) from 11kHz to 22kHz to sense locations. The frequency response below 11kHz is not used since it contains mostly background noise of human activities [25]. Even though this selection makes the sensing of EchoTag audible to humans, the impact of this selection is minimal because EchoTag triggers acoustic sensing very infrequently, i.e., only when the tilt and the WiFi readings match its database as shown in Fig. 3. Moreover, the annoyance caused by sensing with audible sounds is mitigated by reducing the sound volume (e.g., to 5% of the maximum volume) without degrading sensing accuracy. None of the 32 participants in our usability study considered the EchoTag's emitted sound annoying and 7 of them didn't even notice the existence of emitted sound until they were asked to answer related questions in the post-use survey.

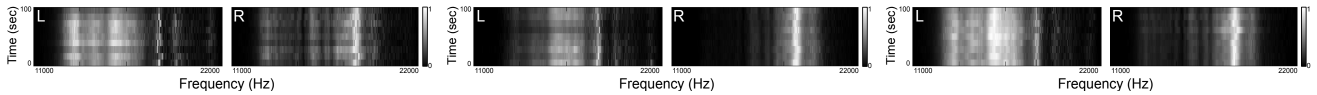
We must also consider the length of the emitted sound, which is correlated with the signal-to-noise-ratio (SNR) of received signals. The longer the sound of a frequency sweep, the more energy at each frequency is collected. However, a long duration of emitted sound introduces a serious problem to the functionality of EchoTag because reflections from far-away objects are collected during this long duration of sensing. Fig. 7 shows the received signal passed by a matched filter, where the peaks indicate received copies of the emitted sound. The first and largest peak in this figure represents the sound directly traveled from the phone's speakers to its microphones and the subsequent peaks represent the reception of environmental reflections. As the purpose of EchoTag is to remember a specific location for future use, it is unnecessary to collect signatures of reflections from far-away objects since those objects are likely to move/change. For example, the object 1.5m away shown in Fig. 7 might be the reflection from the body of a friend

[2] http://developer.android.com/reference/android/media/AudioRecord.html



(a) 100% volume (b) 5% volume (c) 1% volume

Figure 8—Frequency responses at different volumes. Responses of full volume are saturated by sound directly transmitted from speakers while responses at 1% of the maximum volume are too weak to pick up valid features.



(a) 0 sample delay (b) 1 sample delay (c) 2 samples delay

Figure 9—Frequency responses with delay at the right channel. When the emitted sound is intentionally delayed at the right channel, different portions of features are strengthened, which helps enrich the feature space for sensing locations.

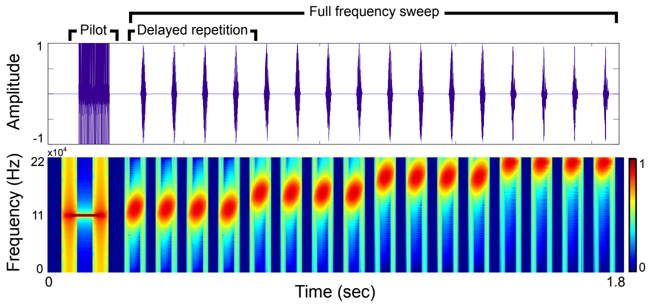


Figure 10—Selected sound signals at EchoTag. The leading pilot is used for time synchronization between speakers and microphones. The following chirps (repeated 4 times each) cover the frequency sweep from 11 to 22kHz. (This figure is scaled for visualization.)

sitting next to the user, and he might move away when EchoTag is triggered to sense locations. One way to mitigate this problem is to truncate the signals generated by the reflections from far-away objects. EchoTag uses 100-sample short signals for sensing and collects only the 200 samples of received signals after the largest peak passes through the matched filter. That is, the sensing range of EchoTag is roughly confined to 1m since the sound of speed is 338m/s and the chosen sample rate is 44.1kHz. The sensing range is actually shown to be shorter than this calculated value since the signals reflected from nearby objects are inherently stronger than those from far-away objects. Our results also confirm that this setting can meet most real-life scenarios in terms of accurate location tagging. One thing to note is that the entire frequency sweep is divided into four smaller 100-sample segments rather than one 100-sample chirp covering the 11–22kHz range. This selection reduces the sample duration (also the sensing range), but keeps enough energy at each frequency for the purpose of sensing locations.

The last parameter in selecting the emitted sound is the time to wait for playing the next chirp after sending a chirp. This parameter is related to the sensing speed of EchoTag. The larger this parameter, the longer the time EchoTag needs for single location sensing. On the other hand, a short wait time causes detection errors since the received signals might accidentally include the reflections of the previously emitted chirp. For example, if EchoTag triggers the next chirp within the 500-th sample shown in Fig. 7, the peaks (i.e., reflections) near the 400-th sample will be added as a noise to the received signals associated with the sensing of the next chirp. From our earlier field study to identify the surrounding objects via sound reflections, we found the speakers and microphones on Galaxy S4 and S5 are able to capture the reflections from objects even 5m away. This phenomenon can also be found in Fig. 7; there is residual energy even after the 1500-th sample. Thus, the interval between two chirps in EchoTag is set to 4500 samples, making its signal sensing time of the entire frequency sweep equal to $4(200 + 4500)/44100 \cong 0.42$ second.

An example of sensing signals is shown in Fig. 10, where a 500-sample pilot is added before the frequency sweep. This pilot is used for synchronization between speakers and microphones because the operating system delays are not consistent in commodity phones. The way EchoTag synchronizes a microphone and a speaker is similar to the sample counting process in BeepBeep [20]. In the current version of EchoTag, this pilot is set as a 11,025Hz tone, which can be improved further by pulse compression [16, 24], but according to our test results, it doesn't make any noticeable difference. Another 10000 samples follow the pilot before the chirp signals are played. Note that 4 chirps in the same frequency range are played consecutively before changing to the next frequency range. This repetition is used to enrich the acoustic feature space as described in the following sections. Current setting of EchoTag makes the total sensing time of EchoTag near 2–3 seconds. After testing EchoTag, most participants of our usability study were satisfied with this latency in sensing locations. Note that in the training phase of EchoTag, each trace is collected with 4 cycles of the above-mentioned frequency sweep to eliminate transient noises, consuming about 10 seconds to collect.

4.3 Volume Control

The volume of an emitted sound plays a critical role in extracting valid signatures from multipath fading. As shown in Fig. 8, when the volume of emitted sound is full (i.e., 100%), a large portion of the feature space is saturated by the sound emitted directly from the phones' speakers. Moreover, emitting sound in full volume makes the sensing process more annoying to the users since EchoTag uses audible frequency ranges. On the other hand, if only 1% of full volume is used to emit sound, the reflections are too weak to be picked up by phones' microphones. Based on our preliminary experiments, setting the phone volume at 5% is found optimal for Galaxy S5. Even though this setting varies from one phone type to another, calibration is needed only once to find the optimal setting.

4.4 Acoustic Signature Enrichment

The goal of EchoTag is to enable accurate location sensing with fine resolution, but according to our experimental results, one shot of the frequency sweep between 11 and 22kHz can distinguish 1cm apart objects with only 75% accuracy. One way to enrich the feature space is repeating the emitted sound which can be used to eliminate the interference caused by transient background noise. Instead of only repeating the emitted sound, EchoTag also adds delay of emitted sound in the right channel at each repetition. This intentional delay at the right channel is designed for two purposes. First, when there are stereo speakers in the phone, such as HTC M8 and Sony Z3, this intentional delay yields an effect similar to beamforming in wireless transmission, which helps us focus on the response in one specific direction at each repetition. We validated this feature to enrich the collected signatures by HTC M8's two front-faced speakers. A similar concept was also adopted in acoustic imaging [13], but EchoTag doesn't need calibration among speakers because the purpose of this delay is used to enrich the feature space rather than pointing to a pre-defined target direction. Second, the intentional delay also helps strengthen features at certain frequencies even when there is only one speaker in the phone. The effects of this delay at the right channel are shown in Fig. 9, where different portions of features are *highlighted* with different delays. Based on the results in Section 8, 4 repetitions with 1 sample delay at the right channel improve EchoTag's sensing accuracy from 75% to 98% in sensing 11 tags, each of which is 1cm apart from its neighboring tags. This way to enrich the acoustic signature space is a unique feature of EchoTag, as it *actively* emits sound to sense the environment, rather than passively collecting existing features.

4.5 Modeling of Sensing Resolution

Suppose two (drawn) tags are at distance d from each other, the sensing signal wavelength is λ, and the angle from one of the nearby objects toward these two tags is θ. The difference of the reflection's travel distance from this object to the two tagged locations is $\delta = 2d * cos\ \theta$. Since the sensing signal wavelength is λ, a change, $\delta > \lambda \not{2}$, in any reflection's travel distance will cause the summation of all acoustic reflections to vary from constructive to destructive combing (or vice versa), thus resulting in a significant change in the acoustic signature. So, if θ of all nearby objects is not close to 0 (which is also rare in the real world), tags separated by more than $\lambda/4$ are to be distinguished by their acoustic signatures. Based on this model, the current setting of EchoTag is capable of identifying tags with a 1cm resolution. Our evaluation also validates this property as shown in the following sections. However, this fine resolution also implies that users should place their phones close enough to the trained location for collecting valid signatures; this is also the reason why EchoTag requires "drawn" tags to remind users where to place their phones. In our usability study, most users didn't have any difficulty in placing phones back at the trained locations with this resolution to activate the tagged functionality of EchoTag. The limitations and future direction of this resolution setting will be discussed further in Section 10.

5. CLASSIFIER

Several classifiers, such as k-nearest neighbors (KNN) and support vector machine (SVM), have been tried for location sensing based on acoustic signatures. Our experimental results show that one-against-all SVM [27] performs best in classifying locations. For example, in the experiment of sensing eleven 1cm apart tags based on the training data collected 30min earlier, 98% accuracy can be achieved by SVM while only 65% test data can be correctly classified via KNN with the Euclidean distance and $k = 5$. We believe this inaccuracy is caused by the small training data size and nonlinear nature of acoustics signatures. For example, a 5mm position change might cause more significant feature changes (in the frequency domain measured by the Euclidean distance) than a 10mm position change since the acoustic signatures capture the superposition of all reflections.

In the one-against-all SVM, n classifiers are trained if there are n tags to sense. A location is classified as the tag k if the k-th classifier outputs the highest prediction probability for that tag. In our test with 2-fold cross validation, the linear kernel achieves the optimal performance. As the results shown in Section 8, the difference of prediction probability between the classifiers trained at the target location and the other locations is greater than 0.5 in most cases, which is adequate for EchoTag to distinguish locations using acoustic signatures.

6. PERFORMANCE OPTIMIZATION

Even though activating microphones and speakers is shown to be more economical than image or WiFi sensing [5], the cost of continuously collecting acoustic signals is still non-negligible and unnecessary. Our measurements with Monsoon Power Monitor [1] on Galaxy S5 show that acoustic sensing consumes 800mW. Moreover, due to the constraints in existing phone microphones, the signals we used are still audible, especially for the pieces of frequency sweep close to 10kHz. The strength of acoustic signal is reduced greatly by lowering the volume, but its continuous use still annoys the users and consumes energy. We have therefore performed further optimizations by avoiding unnecessary acoustic sensing to reduce the power consumption and user annoyance. For example, in terms of EchoTag's functionality, it is useless to sense the environment via acoustic signals when the user keeps the phone in his pocket while walking. This situation can be easily detected by reading inertial sensors. As shown in Fig. 3, EchoTag first checks the status of the surrounding WiFi to ensure that the phone is located in the same target room. Then, the inertial sensor data, such as the accelerometer readings, are used to check if the phone is placed with the same angle as recorded in the database. If both the WiFi status and the inertial readings match the recorded data, one shot of acoustic sensing will be activated to collect the surrounding acoustic signature. The next round of acoustic sensing can be executed only when EchoTag finds the phones moved and the recorded WiFi beacons and inertial readings match. Note that WiFi sensing in EchoTag incurs minimal overhead, since it only needs connected WiFi SSID, which can be directly retrieved from the data already scanned by Android itself via WifiManager. In the current implementation of EchoTag, tilt monitoring only consumes additional 73mW in the background and the power usage of WiFi is negligible since EchoTag uses only the (already) scanned results. We will later evaluate the false trigger rate of this design.

7. IMPLEMENTATION

We implemented EchoTag as an Android background service. Since EchoTag relies only on sensors commonly available in smartphones, it can be readily implemented on other platforms like iOS or Windows. Acquisitions of active acoustic signatures, tilts, and WiFi signals are implemented with Android API while the classifier relying on LIBSVM [6] is written in C layered by Java Native Interface (JNI). The function to trigger applications, such as setting silent mode or playing music, is implemented via Android Intent class. A prototype of EchoTag is also implemented in iOS with classifiers trained in Matlab. In our current implementation, one round of acoustic sensing (including SVM processing) takes about 4 seconds. Most participants in our usability study are satisfied with the current setting.

8. PERFORMANCE EVALUATION

We have conducted a series of experiments to evaluate the performance of EchoTag using Galaxy S5 for two representative scenarios shown in Fig. 11. Certain experiments are also repeated on Galaxy S4/Note3, and iPhone4/5s, but the results are omitted due to space limitation. The first test environment is a lab/office and

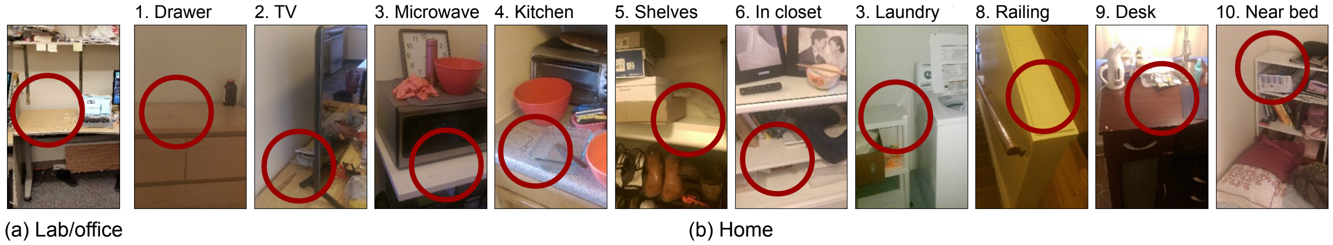


Figure 11—Experiments scenarios. Red circles represent the target location to draw (echo) tags.

the second is a two-floor home. The red circles in Fig. 11 represent the test locations to draw tags. Both scenarios represent real-world settings since people usually work or live in either of these two environments. In the lab/office environment, traces were collected while lab members were chatting and passing through the test locations. During the experiment, an active user kept on working (changing the laptop position and having lunch) on the same table. There are two residents living in the home environment, and one of them is unaware of the purpose of our experiments. The residents behave normally, cooking in the kitchen, watching TV, and cleaning rooms. Thus, our evaluation results that include the interference due to human activities should be representative of real-life usage of EchoTag.

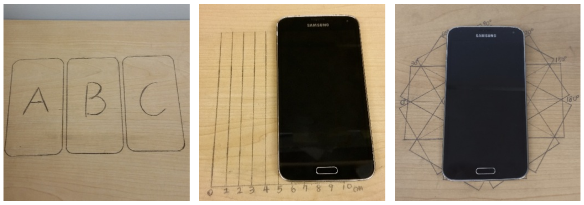
In the lab/office environment, three tag systems shown in Fig. 12 are used to evaluate the *sensing resolution* which is defined as the minimum necessary separation between two tags. The first system is composed of three tags 10cm apart: A, B, and C, as shown in Fig. 12(a). This setting is used to evaluate the basic resolution to support applications of EchoTag (e.g., automatically setting phones to silent mode). The second and third systems include 11 tags which are 1cm (30°) apart from each other as shown in Fig. 12(b) (Fig. 12(c)); this is used to evaluate the maximum sensing resolution of EchoTag. In the home environment, 10 locations are selected as shown in Fig. 11(b). At each location, we marked two tags, A and B, similar to the setting in Fig. 12(a).

In both scenarios, traces are collected at different sampling frequencies and time spans, generating three datasets: 1) 30min, 2) 1day, and 3) 1week. In the 30min dataset, traces are collected every 5 minutes for 30 minutes, which is used to evaluate the baseline performance of EchoTag without considering the environment changes over a long term. The 1day dataset is composed of traces collected every 30 minutes during 10am – 10pm in a day. The purpose of this dataset is to evaluate the performance changes of EchoTag in a single day, which is important for certain applications, such as the silent-mode tag near the bed where the users place their phones every night. The last 1week dataset is collected over a week, which is used to prove the consistency of active acoustic signatures over a one-week period. In the lab/office environment, the 1week dataset is sampled during 9:00am – 9:00pm every day while the 1week dataset in the home environment is tested at 11:00pm.

We believe these experiments can validate the effectiveness of EchoTag in real world. Larger-scale experiments including more users and spanning longer periods are part of our future plan after deploying EchoTag.

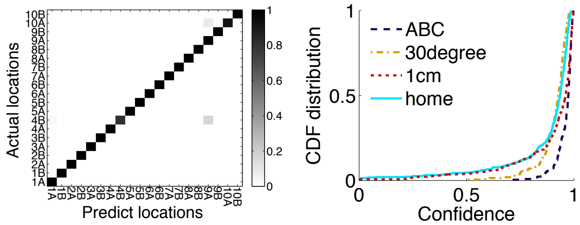
8.1 Accuracy and Resolution

Sensing resolution in EchoTag is defined as the minimum necessary distance/degree between tags, which is an important metric since it is related to the number of tags that can exist in the same environment. On the other hand, the sensing accuracy at a location is defined as the percentage of correct predictions at that location, whereas the overall accuracy is defined as the average of accuracy at all locations. In the lab/office environment of 30min dataset, the average sensing accuracy under all settings is higher than 98%. Orientation changes can be detected by EchoTag since the microphones/speakers are not placed in the middle of the phone, and hence the relative position changes when the phone is rotated. EchoTag can also distinguish 20 tags in a home environment with 95% accuracy. The resulting confusion matrix is shown in Fig. 13(a). This evaluation based on the 30min dataset validates that EchoTag can achieve a sensing resolution of 1cm and at least 30°. Without using any infrastructure, this sensing resolution is the finest among existing methods to differentiate locations. Our measurements show that WiFi RSSI or background acoustic noise can only distinguish the 20 tags at home with 30% accuracy.



(a) ABC tags (b) 1cm tags (c) 30° tags

Figure 12—Tag systems. The first tag system consists of disjoint (echo) tags while the second and third tag systems are composed of overlapped tags 1cm or 30° apart.

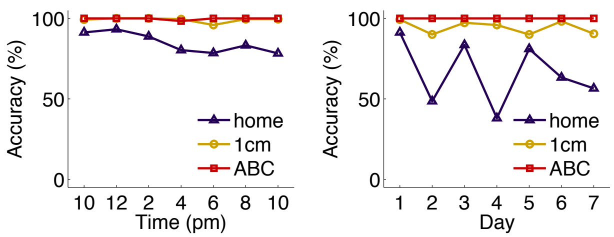


(a) Confusion matrix in home (b) Confidence distributions

Figure 13—Result of 30min dataset. *Confidence* is defined as the prediction probability at the target location minus the largest prediction probability at the other locations.

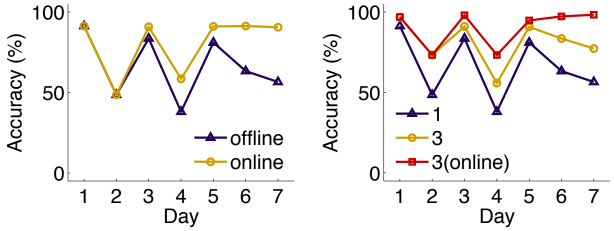
8.2 Uniqueness and Confidence of Acoustic Signature

To measure the uniqueness of acoustic signature, we define a metric called *confidence* as the prediction probability of the classifier trained at the target location minus the largest prediction probability of classifiers at other locations. A high confidence means high feature uniqueness among locations since SVM gets less confused among locations. A prediction is wrong whenever the confidence is less than 0 because SVM will choose another location with the highest prediction probability as the answer. Fig. 13(b) shows the confidence distribution of 30min dataset in all environments. ABC tags get the highest confidence since the tags are separated by more than 10cm and only 3 tags are considered. However, even 20 tags are set at home or tags in office are separated by only 1cm (overlapped), acoustic signatures are still distinct enough to differ-



(a) 1day dataset (b) 1week dataset

Figure 14—Accuracy variation over time/day. Prediction is based on 6 traces collected during the first hour/day.



(a) Online SVM (b) Multiple candidates

Figure 15—Performance of online SVM and providing multiple candidates in the 1week home dataset. Online SVM classifiers are updated using the traces collected in previous days while the traces collected on the same day are excluded.

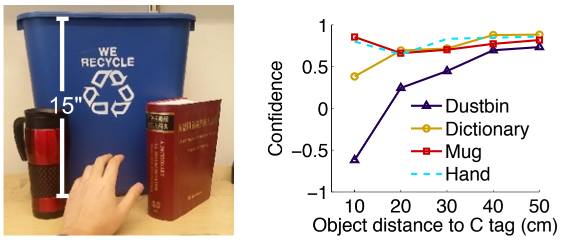
entiate 90% of cases with confidence greater than 0.5. This example demonstrates that the uniqueness of active acoustic signature is good enough to support the function of EchoTag.

8.3 False Positives

The above-mentioned accuracy represents the true positive rate to differentiate locations. To prevent EchoTag from falsely classifying locations without tags as tagged ones, two more traces are recorded on the same surface but 10cm away from each tag. These traces are used to build an additional *No Tag* SVM classifier which determines if the sensed traces belong to one of tagged locations or not. We set 0.5 as the threshold that any sensed location is classified as *No Tag* when the prediction probability of this classifier is greater than 0.5. In the 30min dataset of home environment, the probability to classify random locations without tags as tagged locations (i.e., false positive rate) is only 3%, and this setting only causes a 1% false negative rate. We conducted another test by deploying the ABC tags on three users' office desks for three days. The users carry their phones as usual but are asked not to place their phones inside the drawn tags. In this test, merely 22 acoustic sensings are triggered per day and only 5% of them are falsely classified as being placed at tags with online-updated classifiers. This rate can be reduced further by fusing other signatures which is part of our future work. We also implemented a manual mode (without any false trigger) in which users can press the home button to manually activate acoustic sensing. This manual mode is a reasonable design choice since it is similar to the way Apple Siri or Google Search is triggered.

8.4 Temporal Variation

The purpose of this evaluation is to test how active acoustic signatures change over time. This is an important metric since EchoTag is designed to be able to remember the location at least for a certain period of time. To evaluate this, the average accuracy among tags of 1day and 1week datasets based on the first 6 traces collected in the first hour or day is shown in Fig. 14. As shown in Fig. 14(a), the decay of active acoustic signatures in one day is not significant. In the lab/office environment, the accuracy drops only by 5% in predicting the traces collected 12 hours later. Even one week later, the average accuracy of EchoTag in the lab/office environment is still higher than 90%. However, as shown in Fig. 14, the average accuracy of traces collected in the home environment drops by 15% after 12 hours and the traces collected a week later achieve only 56% accuracy. This phenomenon is caused by a series of environment changes at certain locations. For example, the accuracy drops mainly at 4 locations: (1) drawer, (4) kitchen, (9) desk, and (10) near the bed, which suffer environment changes due to human activities like cooking in the kitchen or taking objects out of drawers. When the above-mentioned objects are excluded from dataset, EchoTag can sense the remaining 12 tags at 6 locations with 82% accuracy even a week later. This result suggests where to put tags is critical to EchoTag's performance. When we consider the tags in the kitchen as an example, if the tags are not placed at the original location near the cooker and stove but on one of the shelves, the acoustic signatures decay as slowly as at other locations. Providing guidelines for where to put tags is part of our future work.



(a) Addition of objects (b) Prediction confidence

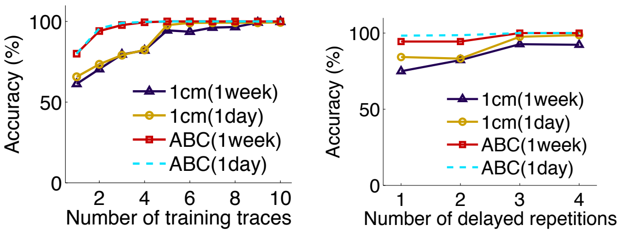
Figure 16—Test of environmental changes. EchoTag gets less confident when the size of an added object is larger and its position is closer to the test locations.

Sensing accuracy over a long term can be improved further in two ways. The first is to use online SVM training. We simulated online SVM by retraining the offline SVM classifiers with the traces collected before the test day (i.e., excluding the same day test data). The cost of retraining classifiers can be further optimized by online SVM [26]. As the results shown in Fig. 15(a), with online training, the average accuracy for the home environment in one week can be increased to 91.5% during the last three days. This online training is possible in real world because we assume users will provide feedback, such as selecting the right application/functions when a wrong prediction is made or suspending wrongly-triggered applications. By monitoring this user reaction after each prediction, online training data can be collected during the normal use of EchoTag. Moreover, in our experiments, only 8.5% of error predictions need this user interaction.

Another way to improve the sensing accuracy over a long term is to provide more than one predicted candidate for users. The candidates in EchoTag are provided based on the prediction probability for each classifier. As shown in Fig. 15(b), when the first three predicted candidates are provided, the accuracy during the last day based only on the first day trace is increased to 77%. Moreover, providing 3 candidates with online SVM training boosts the accuracy of EchoTag to 98% during the last day. Evaluating the overhead of users' interactions with online training feedback and multiple candidates is part of our future work.

8.5 Environmental Disturbances

Similar to the signature decay due to significant environmental changes in the kitchen, we investigate the performance of EchoTag



(a) Training traces (b) Delayed repetitions

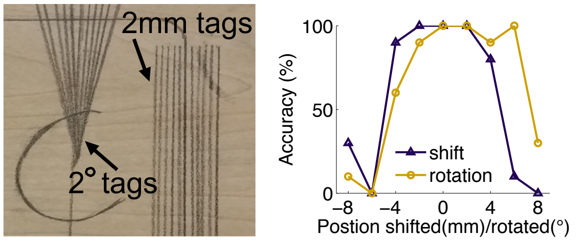
Figure 17—Impact of acoustic feature space. Accuracy is higher than 95% when 5 traces with 4 delayed repetitions are collected.

when objects near a tagged location are different from those during the training. Fig. 16(a) shows 4 selected objects: 1) dustbin, 2) dictionary, 3) mug, and 4) human hands. We add these objects sequentially near the ABC tags and vary their distance to the C tag. The corresponding prediction confidence is used to measure the change of EchoTag's performance. As shown in Fig. 16(b), human hands and small objects like a mug cause little disturbance to EchoTag even when those objects are only 10cm away from the test locations. Medium-size objects like a thick dictionary degrade EchoTag's confidence to 0.38 when it is close to the test locations, but most predictions still remain correct. Placing large objects like a 15" high dustbin around the test locations change the acoustic signatures significantly since it generates lots of strong acoustic reflections. Most predictions are wrong (i.e., confidence < 0) when the dustbin is placed 10cm away from the test locations. It is also the reason why the accuracy in the kitchen degrades after the position of a large cooker is changed. However, this large environment change is less common in real life. For example, users may change their mugs or hands frequently but less likely to move large objects on their office desk.

When a large environmental change occurs, EchoTag needs to retrain its classifier to account for this change. One interesting finding from our 1week trace is that the prediction accuracy in the home environment increased back to 90% after three day classifier online updates. This demonstrates that with enough training data, EchoTag is able to keep only *invariant* features. In future we plan to derive a guideline for setting up tags and providing online feedback when EchoTag finds more training necessary for certain locations. Another interesting finding is that the (dis) appearance of human causes only limited effect on EchoTag because human body is prone to absorb sound signals rather than reflect them. During experiments, a human body (the tester) continuously changed his relative position to the test locations, but no significant performance degradation was observed in spite of the large human body.

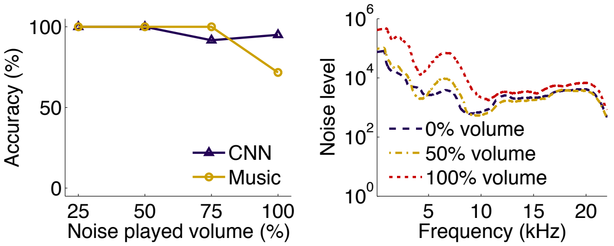
8.6 Acoustic Feature Space

Here we discuss the effect of the feature space selected for sensing locations. We first examine accuracy with different training data sizes. The training data size is relevant to the usability of EchoTag since it represents the time required for users to set tags and the number of features necessary to remember a single location. As shown in Fig. 17(a), in the lab/office scenario, 2–4 traces are able to distinguish rough locations and only 5 traces can achieve 95% accuracy with 1cm resolution. Considering the tradeoff between training overhead and sensing accuracy, the current version of EchoTag sets the size of training data at a single location to 4. In our usability study, these 4 traces at each tag can be collected in 58 seconds on average. We also received a comment from one survey participant that the training process and delay of EchoTag are acceptable since the entire procedure is similar to that of setting up Apple TouchID. Based on our informal experiments with 5 participants on iPhone5s, setting up Apple TouchID required 11–13 training traces with different finger placements, taking about 1 minute. See Section 9 for details of the users' other reactions on this training set size.



(a) Testbed (b) Accuracy of C tag

Figure 18—Tolerance test. Additional tags separated by 2(mm/°) are placed inside the C tag. Test data at C are collected with errors ranging from -8 to 8(mm/°) for knowing the tolerance of EchoTag. Dataset of ABC and 1cm tags are combined, so the accuracy shown is the prediction among 14 locations.



(a) Accuracy at different noise (b) Noise profile of CNN news

Figure 19—Impact of background noise. Predefined noises (i.e., music and CNN news) are played by Macbook Air with different volumes. EchoTag is able to provide effective prediction even when the noise is played at 75% of the full volume.

Next, we study the benefit of our delayed repetitions for sensing locations. The sensing accuracy based on 5 traces with different numbers of delayed repetitions is plotted in Fig. 17(b). As shown in this figure, without help of delayed repetitions to enrich acoustic signatures, the accuracy for the 1cm dataset over a week is only 75% while it can be boosted to 91% when 4 delayed repetitions are used. This way of enriching the feature space is only available in active acoustic sensing since we have the full control of emitted/collected signals. We also find similar effectiveness of this delayed repetitions on phones with stereo speakers like HTC M8, but these results are omitted due to space limitation. The current version of EchoTag uses 4 delayed repetitions.

8.7 Tolerance Range

Since the phone might not be placed exactly at the same locations as it had been trained, it is important to know the maximum tolerance range of EchoTag and if users can handle the tolerance well. To evaluate the maximum tolerance of EchoTag, additional fine-grained tags are drawn inside the C tag as shown in Fig. 18(a). These fine-grained tags are separated by 2(mm/°). Test data of these inside tags are collected with additional errors (i.e., between ±8mm/°), and a prediction is made by the first 4 traces of 30min dataset in the lab/office environment. In the training phase, ABC and 1cm datasets are combined so the prediction is made for sensing 14 locations. The accuracy of the C tag when test data is collected with additional errors is plotted in Fig. 18(b) where the maximum tolerance range of EchoTag is about ±4mm and ±4°. The

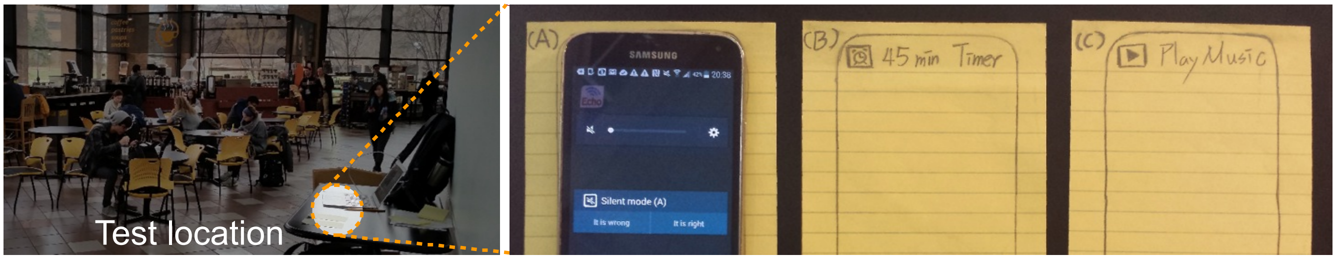


Figure 20—Usability study environments. The test location is selected near the a cafe at a student center. Tags are drawn at memo pads since the table is black. Passers by and students studying in this area are randomly selected to test `EchoTag`.

reason why accuracy with different degree errors is not centered at 0° might be the measurement errors in collecting the training data (i.e., the training data is also collected with a few degree errors). This result also matches our resolution test, where `EchoTag` can provide 1cm resolution since features of tags separated by 8mm vary significantly. With the help of drawn tags, this tolerance range is good enough to allow users to put their phones back at the tags for triggering `EchoTag`. Our usability study of 32 participants validates this hypothesis since most of them think it is easy to place phones on the tags. We will discuss how to enhance the user experience with this limitation of tolerance range in Section 10.

8.8 Noise Robustness

The last issue of `EchoTag` to address is its ability to sense locations in a noisy environment. Basically, the results discussed thus far were obtained in real-life scenarios where traces were collected in the presence of noises from TV, people chatting, and air condition fans. To further test the robustness against noises, we collected traces in a lab/office environment when a laptop was placed 30cm away from tags on the same surface. This laptop is set to either play a song ("I'm yours – Jason Marz") or a clip of CNN news with different volumes. The traces of ABC and the 1cm dataset are combined to make prediction for 14 locations. The results of this experiment are shown in Fig. 19. As shown in Fig. 19(a), `EchoTag` can tolerate the music noise with 75% volume and can perform normally even with noise from CNN news with 100% volume. In our measurements, the intensity of noise from the music with 75% volume and CNN news with 100% volume is about 11*dB* higher than the office background noise. Even though `EchoTag` is unable to work accurately (i.e., only 71% among 14 locations) when the music was played with 100% volume, this is not a typical operation scenario of `EchoTag` since it incurs 17*dB* higher noise.

`EchoTag` is robust to most real-world noises mainly because it relies on signatures actively generated by itself, rather than passively collected from background. That is, as the noise profile of CNN news shown in Fig. 19(b), most noise from human speaking occurs in frequencies less than 10kHz while `EchoTag` uses higher frequencies to sense locations. This phenomenon is also consistent with the results shown in [25]. The current setting of `EchoTag` is already resilient to the noise in our usual environment. Our usability study done at a noisy location near a cafe also validates the robustness against noise. Incorporating other sources of signatures that are free from acoustic noise is part of our future work.

9. Usability Study

In this section, we explore how real users of `EchoTag` perceive its performance. For example, we would like to answer a question like "can users easily place phones on drawn tags?" To evaluate the usability of `EchoTag`, 32 (9 female and 23 male) participants were recruited at our university students' activity center. The test location was the table near a cafe as shown in Fig. 20. Participants are randomly selected from those passing by this area. All participants didn't know `EchoTag` and its developers, but all have experience in using smart phones (11 Androids, 19 iPhones, and 2 others). Most participants are university students of age 20–29 while 6 of them are not. We first introduced the functionality of `EchoTag` to users and its three representative applications as shown in Section 1. Then, three tags (i.e., turning on silent mode, setting a 45min timer, and playing music) were deployed at the table. Since the table surface is black and is the university's property, we drew the tags on sheets of 5×6 paper which were then attached on the table. Our survey results also validate that drawing tags on attached papers will not affect the accuracy of `EchoTag`. After users were informed of the purpose of `EchoTag`, we asked them to put phones on the tags and `EchoTag` is triggered automatically to make its prediction on attached applications. The participants were asked to first put phones inside the tag for measuring prediction accuracy, and then slightly move the phones away from the tags until a wrong prediction is made for testing the tolerance range of `EchoTag`. After trying the functionality of `EchoTag` with the tags we trained, the participants were asked to train their own tags. This process helps users "sense" the amount of effort to set up and use `EchoTag`, indicating the average user training time. The participants then filled out the survey forms. Each study lasted 15–20 min.

Questions	Disagree	No option	Agree
Sensing accuracy is useful	1	0	31
Sensing noise is acceptable	0	3	29
Sensing delay is acceptable	1	6	25
Placing phones inside (echo)tags is easy	0	3	29
`EchoTag` can help me remember turning on silent mode when going to sleep	2	5	25
`EchoTag` can help me remember setting the timer for taking washed clothes	5	3	24
`EchoTag` can save my time in activating apps under specific scenarios	1	0	31

Table 2—Usability survey results of 32 participants.

The main results of this usability study are summarized in Table 2. 31 of 32 participants think the sensing accuracy of `EchoTag` is adequate for its intended applications. During this study, three users experienced a bug in our app. In these cases, we restarted the whole survey process and asked the participants to try `EchoTag` again. The average accuracy including these three buggy traces were 87%, while the results excluding these three cases were 92%. One thing to note is that even those three users experienced the low sensing accuracy during the first trial, still think the overall detection accuracy of `EchoTag` is acceptable and useful. Moreover, 25 of the participants think the prediction delay is acceptable for the purpose of `EchoTag`. In another question to learn their expectation of `EchoTag`, only 2 of users hope to have the sensing delay less than 1 second and 12 users hope the sensing delay can be shortened to about 2 seconds.

Based on participants' experience in placing phones on tags, 29 of them think it is easy to do. Five participants put their phones too far away from the drawn tags during the first trial, but they were able to place their phones on the tags after they were told of it. Only one participant made a comment that he expected a larger tolerance range. Actually, there is an inherent tradeoff between the tolerance range and the training overhead. For example, a very long training sequence that moves phones at all possible locations near the drawn tags can increase the tolerance range, but with a significant overhead in the training phase. In the current version of EchoTag, 4 repetitions are used in the training phase to collect signatures at a single location. Between two consecutive repetitions, the phones need to be taken away from, and then placed back on tags for sensing valid signatures. On average, users need 58 seconds to finish the whole training phase for a tag. 8 participants expected less than 2 training repetitions at each location, while 17 of participants think the 4 repetitions setting is reasonable since it is only one-time training. Considering this trade-off between accuracy, tolerance, and training overhead, the existing setting of EchoTag satisfies most scenarios and users.

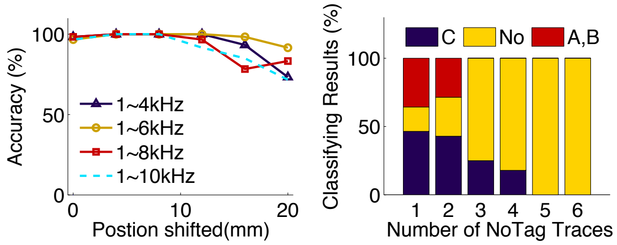
According to the survey related to potential scenarios based on EchoTag, 25 participants agree the silent mode tag can help them remember setting phones to stay quiet during sleep, and 24 of them agree the auto-timer scenario can help them take clothes out of a washing machine. These results indicate that the users see benefits from EchoTag in remembering things to do in specific situations since the functionality of EchoTag provides natural connections between the locations and things-to-do. Moreover, 31 participants also think the automation of triggered applications can save time in finding and launching desired applications.

10. Discussion and Limitations

The purpose of EchoTag is to provide a novel way of accurately tagging locations. With the realization of this functionality, we believe many advanced applications can be realized. Most participants in our usability study agree with the scenarios we considered, and they also provide other possible uses of EchoTag.

10.1 Potential Applications

Similar to those applications mentioned so far, participants also suggest use of an auto-set timer during cooking or auto-set silent mode in an office environment. These applications can be classified into two categories based on the resulting benefits: 1) helping users remember things, and 2) saving time for users in finding and launching desired applications. For example, applications like auto-set silent mode do not save a significant amount of time since pressing a phone's power button can also set up silent mode manually. However, this function turns out to receive most support from the participants because it is easy to forget setting silent mode before going to bed everyday. Instead of doing it manually, a natural combination of a location (i.e., a tag near bed) and target applications/functions (i.e., silent mode) helps users remember them more easily. Other suggested applications like automatically turning off lights or activating a special app for monitoring the sleep behavior are also in this category. On the other hand, the applications like automatically playing favorite songs (movies) on speakers (TV) help users find the expected applications/functions with less time due to the hint of tagged locations. For example, we received a feedback saying that EchoTag can be used to set up Google Map and activate the car mode when a tag is placed inside the car (e.g., an iPhone stand). Actually, this task can be easily remembered whenever the user gets into his car, but it is time-consuming to find and enable these applications. In this scenario, the natural combination of locations and applications can help users *filter* out unnecessary information and save their time in triggering the desired applications.



(a) Tolerance in lower frequency (b) NoTag SVM classifiers

Figure 21—Extension of tolerance range. The tolerance range can be extended by sensing tags with lower-frequency signals. Building 'NoTag' classifiers can also prevent EchoTag from incorrect classification of misplacements.

10.2 Limitation of Tolerance Range

As shown in Section 8, the merit of EchoTag's fine resolution comes with the limitation that the tolerance range of current setting is about 0.4cm. Even though most participants in our user study were able to place phones at the tagged locations accurately after they were instructed on how to use EchoTag, this fine resolution and its limited tolerance range may not be needed for certain applications and cause unintended prediction errors when the phone is not placed correctly. To meet such applications needs, one can take the following two approaches.

First, we can enlarge the tolerance range of EchoTag (while decreasing its sensing resolution). Based on the mathematical model in Section 4.5, we can lower the sensing frequency to extend the tolerance range. For example, setting the highest sensing frequency to 6kHz can increase the tolerance range from 0.4cm to about 1.4cm. This increase of tolerance range is plotted in Fig. 21(a), showing that lower-frequency signals are better in identifying tags with large placement errors. However, lowering sensing frequency will increase the audibility of sensing signal and decrease the feature space. For example, sensing with 1∼4kHz lowers overall accuracy because the collected signature is not enough to be distinguished from others. Studying the tradeoff between tolerance range and other concerns is part of our future work.

Second, instead of trying hard to classify misplacements as trained tags, EchoTag can simply opt to report "There is no tag" once the location is out of the tag's tolerance range. We choose to use the same 'NoTag SVM' classifier as introduced in Section 8.3, which is built from traces in the same surface but at least 0.4cm away from the tags. That is, EchoTag identified locations as "There is no tag" if the trace gets the prediction probability of NoTag classifier which is greater than 0.3 and also larger than the prediction probability of other classifiers. With this method, if the user places his phone outside of the operation range of EchoTag, the message of "There is no tag" (rather than identifying a wrong tag and then triggering a wrong function) will tell the user he needs to place the phone more precisely to activate the tagged function. The result of the same experiment of identifying C tag with traces collected 0.4cm away from the trained location is shown in Fig. 21(b). Without the help of 'NoTag' classifier, 50% of those misplaced traces around the C tag were classified as wrong tags, i.e., as A and B. However, with enough NoTag training traces, EchoTag can identify the misplacements with a high probability, thus not triggering functions attached to wrong tags. Note that NoTag SVM with 6 training traces in this experiment only causes 1% false negatives when the trace is collected at the right location (within the tolerance range), which also matches the result reported in Section 8.3.

11. Conclusion

We have proposed and implemented `EchoTag`, a novel indoor location tagging based only on built-in sensors available in commodity phones. `EchoTag` is designed to be able to remember indoor locations with 1cm resolution, enabling the realization of many new applications. The main idea of `EchoTag` is to *actively* sense the environments via acoustic signatures. With the help of active sensing, a fine-grained control of collected signatures can be achieved for either enriching the feature space or removing environmental interferences. Our evaluation in different environments validates the capability of `EchoTag` to meet the users' need. In future, we would like collect larger datasets after deploying `EchoTag`.

Acknowledgements

The authors would like to thank Kassem Fawaz, Arun Ganesan, Huan Feng, Dongyao Chen, Yuanchao Shu, Tsung-Han Lin, Lung-Pan Cheng, Yu-Ting Hsu, Chun-Hung Hsiao, Ye-Sheng Kuo, the anonymous reviewers and the shepherd for constructive comments on the earlier versions of this paper.

12. References

[1] Monsoon Power Monitor. `http://www.msoon.com/LabEquipment/PowerMonitor/`.

[2] Near Field Communication. `http://www.nfc-forum.org`.

[3] Azizyan, M., Constandache, I., and Roy Choudhury, R. Surroundsense: Mobile phone localization via ambience fingerprinting. In *Proceedings of ACM MobiCom '09*, pp. 261–272.

[4] Bahl, P., and Padmanabhan, V. Radar: an in-building RF-based user location and tracking system. In *Proceedings of IEEE INFOCOM '00*, pp. 775–784 vol.2.

[5] Ben Abdesslem, F., Phillips, A., and Henderson, T. Less is more: Energy-efficient mobile sensing with senseless. In *Proceedings of the ACM MobiHeld '09*, pp. 61–62.

[6] Chang, C.-C., and Lin, C.-J. LIBSVM: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology 2* (2011), 27:1–27:27. Software available at `http://www.csie.ntu.edu.tw/~cjlin/libsvm`.

[7] Chen, Y., Lymberopoulos, D., Liu, J., and Priyantha, B. FM-based indoor localization. In *Proceedings of ACM MobiSys '12*, pp. 169–182.

[8] Chung, J., Donahoe, M., Schmandt, C., Kim, I.-J., Razavai, P., and Wiseman, M. Indoor location sensing using geo-magnetism. In *Proceedings of ACM MobiSys '11*, pp. 141–154.

[9] Goel, M., Lee, B., Islam Aumi, M. T., Patel, S., Borriello, G., Hibino, S., and Begole, B. Surfacelink: Using inertial and acoustic sensing to enable multi-device interaction on a surface. In *Proceedings of ACM CHI '14*, pp. 1387–1396.

[10] Gustafsson, T., Rao, B., and Trivedi, M. Source localization in reverberant environments: modeling and statistical analysis. *Speech and Audio Processing, IEEE Transactions on* (2003), 791–803.

[11] Harrison, C., Schwarz, J., and Hudson, S. E. Tapsense: Enhancing finger interaction on touch surfaces. In *Proceedings of ACM UIST '11*, pp. 627–636.

[12] Harrison, C., Tan, D., and Morris, D. Skinput: Appropriating the body as an input surface. In *Proceedings of ACM CHI '10*, pp. 453–462.

[13] Izquierdo-Fuente, A., del Val, L., Jiménez, M. I., and Villacorta, J. J. Performance evaluation of a biometric system based on acoustic images. In *Sensors 2011*, vol. 11, Molecular Diversity Preservation International, pp. 9499–9519.

[14] Kunze, K., and Lukowicz, P. Symbolic object localization through active sampling of acceleration and sound signatures. In *Proceedings of UbiComp '07*, pp. 163–180.

[15] Kuo, Y.-S., Pannuto, P., Hsiao, K.-J., and Dutta, P. Luxapose: Indoor positioning with mobile phones and visible light. In *Proceedings of ACM MobiCom '14*, pp. 447–458.

[16] Lazik, P., and Rowe, A. Indoor pseudo-ranging of mobile devices using ultrasonic chirps. In *Proceedings of ACM SenSys '12*, pp. 391–392.

[17] Liu, K., Liu, X., and Li, X. Guoguo: Enabling fine-grained indoor localization via smartphone. In *Proceeding of ACM MobiSys '13*, pp. 235–248.

[18] Nirjon, S., Dickerson, R. F., Asare, P., Li, Q., Hong, D., Stankovic, J. A., Hu, P., Shen, G., and Jiang, X. Auditeur: A mobile-cloud service platform for acoustic event detection on smartphones. In *Proceeding of ACM MobiSys '13*, pp. 403–416.

[19] Ono, M., Shizuki, B., and Tanaka, J. Touch & activate: Adding interactivity to existing objects using active acoustic sensing. In *Proceedings of ACM UIST '13*, pp. 31–40.

[20] Peng, C., Shen, G., Zhang, Y., Li, Y., and Tan, K. Beepbeep: A high accuracy acoustic ranging system using cots mobile devices. In *Proceedings of ACM SenSys '07*, pp. 1–14.

[21] Plack, C. J. The sense of hearing. Lawrence Erlbaum Associates, Inc.

[22] Priyantha, N. B., Chakraborty, A., and Balakrishnan, H. The cricket location-support system. In *Proceedings of ACM MobiCom '00*, pp. 32–43.

[23] Rossi, M., Seiter, J., Amft, O., Buchmeier, S., and Tröster, G. Roomsense: An indoor positioning system for smartphones using active sound probing. In *Proceedings of the 4th Augmented Human International Conference* (2013), pp. 89–95.

[24] Salemian, S., Jamshihi, M., and Rafiee, A. Radar pulse compression techniques. In *Proceedings of WSEAS AEE'05*, pp. 203–209.

[25] Tarzia, S. P., Dinda, P. A., Dick, R. P., and Memik, G. Indoor localization without infrastructure using the acoustic background spectrum. In *Proceedings of ACM MobiSys '11*, pp. 155–168.

[26] Tax, D., and Laskov, P. Online svm learning: from classification to data description and back. In *IEEE Neural Networks for Signal Processing, 2003*, pp. 499–508.

[27] Vapnik, V. N. *Statistical Learning Theory*. Wiley-Interscience, 1998.

[28] Wang, J., Zhao, K., Zhang, X., and Peng, C. Ubiquitous keyboard for small mobile devices: Harnessing multipath fading for fine-grained keystroke localization. In *Proceedings of ACM MobiSys '14*, pp. 14–27.

[29] Xiong, J., and Jamieson, K. Towards fine-grained radio-based indoor location. In *Proceedings of ACM HotMobile '12*, pp. 13:1–13:6.

[30] Youssef, M., and Agrawala, A. The horus wlan location determination system. In *Proceedings of ACM MobiSys '05*, pp. 205–218.

[31] Zhou, Z., Diao, W., Liu, X., and Zhang, K. Acoustic fingerprinting revisited: Generate stable device id stealthily with inaudible sound. In *Proceedings of ACM CCS '14*, pp. 429–440.