Public Restroom Detection on Mobile Phone via Active Probing

Mingming Fan, Alexander Travis Adams, Khai N. Truong
Department of Software and Information Systems
University of North Carolina, Charlotte
{mfan, aadams85, ktruong8}@uncc.edu

ABSTRACT
Although there are clear benefits to automatic image capture services by wearable devices, image capture sometimes happens in sensitive spaces where camera use is not appropriate. In this paper, we tackle this problem by focusing on detecting when the user of a wearable device is located in a specific type of private space—the public restroom—so that the image capture can be disabled. We present an infrastructure-independent method that uses just the microphone and the speaker on a commodity mobile phone. Our method actively probes the environment by playing a 0.1 seconds sine wave sweep sound and then analyzes the impulse response (IR) by extracting MFCCs features. These features are then used to train an SVM model. Our evaluation results show that we can train a general restroom model which is able to recognize new restrooms. We demonstrate that this approach works on different phone hardware. Furthermore, the volume levels, occupancy and presence of other sounds do not affect recognition in significant ways. We discuss three types of errors that the prediction model has and evaluate two proposed smoothing algorithms for improving recognition.

Author Keywords
Restroom detection, room impulse response, active probing, pattern recognition

ACM Classification Keywords
H.5.5. [Sound and music computing]: Signal analysis, synthesis, and processing

INTRODUCTION
Wearable cameras produce personal image-based records which can be used in a variety of ways. For example, researchers have used such records to investigate health behaviors (such as exercise and diet [9, 10]), help people with memory loss recall past events [8], increase parental understanding of the needs of children with autism [15, 18], and improve everyday memory and social skills for children with disabilities [1]. Although there are demonstrated benefits to wearing an always-on and automatically recording personal camera, there are also documented concerns of recording others, particularly in sensitive spaces [1, 3, 4, 9, 10, 11]. As a result, many researchers and users have expressed a need for a mechanism to temporarily disable capture. However, even when manual *"privacy buttons exist and wearable cameras can be removed, it is not uncommon for participants to report that they forgot they were wearing the unit. Therefore, the participant inadvertently might collect inappropriate images, such as going to the bathroom"* [11]. *"With thousands of images automatically recorded every day, … [the user] only deletes unwanted images if he comes across them, as searching for them would take too much time"* [4]. Therefore, how to turn off wearable cameras automatically in sensitive or private spaces is an important research problem.

We tackle this problem by exploring how to detect a specific type of private space where image recording is socially inappropriate—the public restroom. Many researchers have identified the restroom as a specific type of space where they want to suspend capture (*e.g.,* [1, 3, 4, 11]). We focus on public restrooms, in particular, because of the potential for others to be recorded in the captured images there. This problem is challenging for two reasons. First, infrastructure-dependent indoor localization approaches (*e.g.,* cellular, WiFi, and visible light) depend on the infrastructure coverage and floor maps to identify a restroom's location. Infrastructure-independent indoor localization approaches (*e.g.,* inertial sensors on phone) would still require floor maps in order to reason and determine if the user's location is inside a restroom. However, such localization methods fail when the user is outside of an infrastructure's coverage or at a location where floor maps have not yet been developed. Alternatively, video or image based approaches can be employed to detect restrooms [17, 19, 20, 25, 26]. Unfortunately, vision based techniques can sometimes miss signage located immediately outside the space [26]. These methods still can be used inside the space to detect the presence of objects, construction material, and fixtures commonly found in restrooms to reason that must be where the user is located (*e.g.,* [19]); however, this violates the original motivation of not wanting recording to happen there in the first place. Furthermore, recording must still be on to determine when the user has left the space in order to resume the archiving of captured images.

ISWC'14, September 13 - 17 2014, Seattle, WA, USA

http://dx.doi.org/10.1145/2634317.2634320

Figure 1. (Left) Wearable phone attachment for performing and analyzing impulse response and image capture; (Right) 8 sample restrooms used in our study.

In this paper, we present an infrastructure-independent approach, which uses the hardware already found on commodity mobile phones to emit a probing sound and analyze the impulse response (IR) of a space. Our work uses active probing to detect restrooms (a *type* of space) rather than a *specific* space amongst a defined set of locations from which a room type classifier has been trained. We demonstrate that our model can predict new spaces and that the approach works on different phones. We discuss how the volume level does not affect the model in significant ways. Therefore we can minimize the obtrusiveness of the sweep by outputting it at a lower volume. We also show that our model maintains its prediction performance despite of the occupancy and the presence of other sounds in restrooms. We discuss three types of errors that our SVM model has and propose two smoothing algorithms to improve the prediction accuracy.

THEORY OF OPERATION

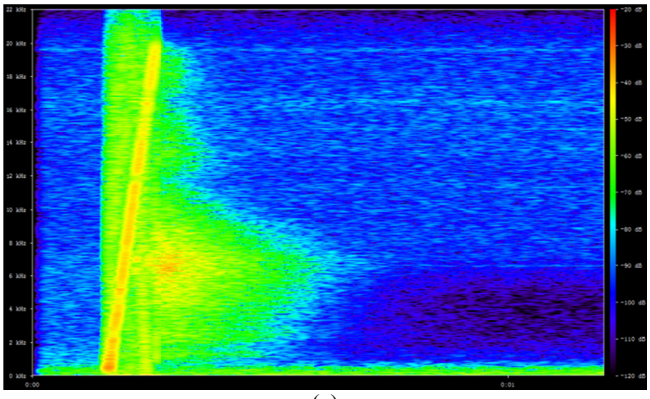
An important and commonly used measurement to analyze and characterize the acoustic properties of different environments and materials is the Impulse Response (IR) [24]. IRs contain time-domain acoustic properties, one of which is the reverberation and its components (*i.e.,* direct sound, decay time, early reflection, and echoes). A common method for capturing the IR of a room is to use a sine wave sweep over a predetermined frequency range to excite the acoustics of a room [14]. The sine wave sweep and a specified amount of time afterwards is recorded and then analyzed to understand the behavior of each audible frequency over time in an environment. IR analysis is commonly used by audio engineers to help them design and tune their systems in order to enhance the sound and avoid any undesired feedback or noise. Acousticians use this technique as an analysis tool to aid them when designing a space. IRs are also used in convolution reverb processing to create a digital representation of the acoustics of different environments.

The acoustic characteristics of an environment are contingent on its dimensions and ability to absorb sound (absorption coefficient). This is a function of several constant factors including the shape, size, construction materials, and objects inside of the construct (*e.g.,* chairs, desks). There are also many variables, such as humidity, and percentage of occupancy, that can change the acoustics continuously over time. These characteristics not only affect reverberation time, but they also affect other parameters of sound such as diffraction, refraction, and reflection [23].

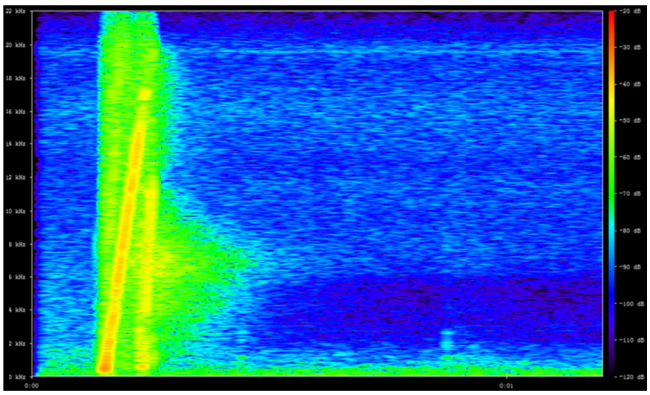
Because no two environments are exactly the same, they all have unique acoustic characteristics or fingerprints. However, many constructs have very similar fingerprints due to its purpose. One particular environment that has unique qualities from other types of environments, yet also has strong correlations between similar environments that serve the same purpose, is the restroom.

In both the public and private space, restrooms have similar affordances that greatly impact the acoustic fingerprint. These affordances include water resistant floors and walls, toilets, and sinks. While public restrooms have stalls and private restrooms have showers/tubs, they both demonstrate similar acoustic responses and can be identified as restrooms from that response (even to the human ear over the phone). This is partially due to the common layout of restrooms but can also be attributed to the materials used on the surfaces and the items found in a restroom (which all have similar absorption coefficients [14]).

Our system leverages the natural acoustic traits of different spaces (in particular the restroom) by exciting the natural acoustics via IR. The IR of different spaces can then be processed to extract acoustic features which are then used to train a classifier to identify the type of space where the user is located. This active probing approach differs from existing computational auditory scene recognition (CASR) methods which extract features and trains classifiers on environment or human generated sounds (such as traffic noise or sink usage) collected from the target scenes [2, 16, 21]. An active probing approach allows for the classification of restrooms to happen even before the user begins to use the space (*e.g.,* urinating, defecating, flushing, and hand washing).



(a)



(b)

Figure 2. (a) Impulse response of a restroom; (b) Impulse response of an office.

Active probing has been explored as a localization approach [13,22]. Kunze and Lukowicz [13] showed how to detect a mobile phone's symbolic location from a set of pre-defined ones by using vibration and short narrow frequency 'beeps'. By emitting a sound and analyzing the impulse response, Rossi *et al*. [22] were able to classify a user's room-level location among 20 rooms at 98% accuracy. In our work, we use a Sine Wave Sweep instead of Maximum Length Sequence (MLS) for active probing, because a sine wave sweep exhibits better tolerance to nonlinearity and time-variance of the probed spaces [6]. This allows us to tackle the goal of recognizing space type.

SYSTEM IMPLEMENTATION

Impulse Response Measurement

To perform the IR measurement, we use the microphone and the speaker already available on commodity mobile phones (see Figure 1). We developed a measurement application that first starts recording with the built-in microphone, and then it outputs a sine wave sweep from 20Hz – 20kHz from the built-in speaker. After the sine wave sweep stops, the application continues to record for one second, allowing it to capture fully the reverberation of the measured space. The measurement application records the IR at 44.1 KHz with 16 bit depth. Figure 2 shows how spectrograms of IRs collected on a Nexus 4 phone differ for a restroom and a non-restroom space.

Stans *et al.* previously compared different impulse response measurements (MLS, Inverse Repeated Sequence (IRS), Time-Stretched Pulses, and Sine Wave Sweep). In this comparison, they found that the sine wave sweep is an ideal choice for non-occupied spaces. It also has the benefit of not requiring any tedious calibrations (which is important in creating a robust mobile system) [24]. Restrooms generally have very few people coming and going, and are often unoccupied. This enables the sine wave sweep to perform optimally, making it the best option to use for IR analysis. Furthermore, a commodity mobile phone's hardware is typically optimized for the range of human hearing, which matches the ideal frequency range for using a sine wave sweep to capture IRs. Because the sweep is audible to humans, we want it to be as short as possible yet still be able to produce a noticeable probing effect. We tested sweeps of 0.01, 0.1 and 1s in duration and found that the 0.1s duration provides the best balance for capturing the acoustics of a space and minimizing the intrusiveness.

Feature Extraction

The measurement application intentionally starts recording before it outputs the sweep and stops recording one second after the sweep has completed to fully capture the effect of an IR. The IR is present in the recording after the measurement application begins to output the sweep. There is a variable latency between when the measurement application requests the OS to output the sine wave sweep and when it actually plays. Due to this delay, the sweep starts at an unknown time point. To estimate the start of the sweep: 1) Divide the IR into frames using a sliding window size W = 1024 samples (23 milliseconds). Each adjacent two windows have 50% overlap with each other. 2) Smooth each window using the Hanning function. 3) Calculate the FFT magnitudes for each window. $FFT_{j,k}$ represents the Kth ($K = 0,\ldots,W\text{-}1$) FFT magnitude of the window j (j = 1…N, N:# of frames in a sweep record). The number of frames S inside the sweep itself is calculated as: $S = \frac{44100 * 0.1}{W * 50\%}$. 4) Estimate the index to the window at the start of the sweep using the following optimization: $arg\ max_{i\in\{1\ldots N\}} \sum_{j=i}^{\min\{i+S,N\}} \sum_{K=0}^{W-1} FFT_{j,k}$.

Processing the rest of the recording after the start of the sweep guarantees that the complete IR is analyzed. However, it may also process extraneous information because the IRs captured in different types of spaces decay at different rates. We have found experimentally that the optimal amount of the recording after the sweep starts to perform feature extraction on is 0.4 s (*0.1 s of the sweep itself + 0.3 s after the sweep stops*). In the evaluation section, we describe the process for determining this value through tests of different durations after the start of sweep.

For the identified optimal amount of the recording to use for feature extraction, we extract the mel-frequency cepstral coefficients (MFCCs) from each window. MFCCs are commonly used as features in the speech recognition and

computational auditory scene recognition [2, 16, 21]. In our implementation, we calculate the first 13 MFCCs by applying 23 Mel filters, remove the first MFCC, and then use the rest to form a feature vector: $(F_{1,i}, \ldots, F_{12,i})$. In the final step, we aggregate all the MFCCs of all the frames by calculating the mean values of each one of the MFCCs using the following equation: $\overline{F_k} = \frac{\sum_{i=1}^{N} F_{k,i}}{N}$ (k=1…12, N: # of frames in the optimal amount of the recording).

Classification

Classifying restrooms *vs.* non-restrooms falls under what is generally referred in machine learning as One Class Classification, because restroom is the target class that we are interested in identifying among all the possible spaces. However, one-class classifiers tend to be conservative in their predictions. We have tried two one-class classification algorithms (Hempstalk *et al.* [7] and the one-class classification in LibSVM [5]), and found that they often predict a new sample as "unknown" and therefore yield low recalls for the restroom class.

Because restrooms are built to serve the same functionality, conform to building codes, use similar construction materials, and have similar materials inside of the construct (*e.g.,* toilets, wash basins), there should be high inner consistency in the restroom class, which at the same time might be highly distinguishable from the "non-restroom" class. Furthermore, because people spend much more time outside of restrooms, we can easily collect a variety of non-restroom data to help the classifier learn the classification boundary. Thus, we decided to treat restroom detection as a binary-class classification, which predicts current room type as either restroom or non-restroom. We leverage LibSVM [5] for classification in our evaluation.

PERFORMANCE EVALUATION

Using a Galaxy Nexus, we collected IR data for 103 public restrooms from 49 different buildings (built between the 1960s – 2006). To help the model learn the classification boundary, we also collected non-restroom data. We attempted to sample as diverse a set of non-restroom spaces as possible (*e.g.,* hallway, elevator, locker room, outdoor, classroom, shop, and bus station). However, we note that it is hardly possible to cover all different types of non-restroom spaces that a user can potentially visit in our evaluation. Information about our collected dataset is shown in the first row of Table 1. Figure 3 shows the types of places where we collected the restroom and non-restroom data.

For each space, we collected 30 samples at different spots perceived to support circulation. Circulation is a term in building architecture to refer to the way that people move through and interact with the space. For instance, in a men's restroom, the typical circulation would be from the door to urinals / toilet stalls, the washstand, the paper towel racks, and then back to the door. While collecting the data, we held the phone still in one hand in front of the chest to

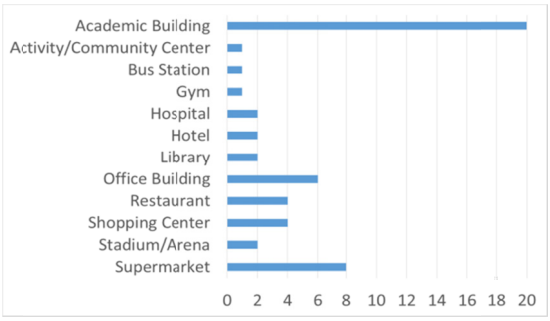


Figure 3. Different types of places where data was collected

simulate wearing the device around the neck. The phone remained stationary while recording the impulse response. We took care to hold the phone such that both the speaker and the microphone were not covered or blocked in anyway.

Optimal Amount of Recording for Feature Extraction

As previously described, the application begins recording slightly before the sweep and continues for a full second. This guarantees that the full IR is captured within the recording. We can remove extraneous information between the start of the recording and the start of the sweep by extracting features from the start of the sweep instead of the start of the recording. To identify the optimal amount of recording after the start of the sweep to use for feature extraction, we selected 10 different durations at 0.1 s interval after the sweep has stopped after the sweep stops (0.1~1 s). For each duration (D = 0.1,…,1.0), we extracted features based on all 7474 recordings by the Galaxy Nexus, and performed a 10-fold cross validation with SVM as classifier. The results are shown in Figure 4. All the five performance measurements indicate that 0.3 seconds duration after the sweep has stopped is the optimal amount of the recording to use for feature extraction. It also shows that models trained with longer durations than 0.3 seconds are less accurate. Therefore, in all following evaluations, we used this finding and performed feature extraction only on the 0.4 seconds portion of the recording after the start of the sweep (*0.1 s sweep itself + 0.3 s duration after the sweep has stopped*).

Efficacy of Model in Classifying New Restroom Spaces

To assess the model's efficacy in classifying new restroom spaces, we evaluated how the model converged as the number of restrooms used for training (training set size). Using the Galaxy Nexus dataset (103 restrooms), we

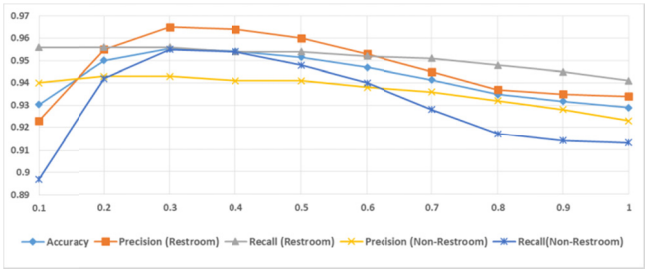


Figure 4. Five measurements of the model using 10 different durations from the start of the sweep for the feature extraction.

gradually increased the number of restrooms (N=1,…,103) used for training. For each number N, the evaluation procedure worked in this manner. First, we randomly chose N restrooms from the 103 restrooms. Second, we randomly chose the same percentage of non-restroom IRs from all non-restroom IRs. Thus, if the total number of restroom and non-restroom are T_{rm}, T_{nrm} , then the number of IR in these N restroom is R_N. Therefore, the number of non-restroom IRs chosen was: $NR_N = T_{nrm} * \frac{R_N}{T_{rm}}$. Third, we used SVM as the classifier and performed a 10-fold cross validation on these $R_N + NR_N$ IRs. Fourth, to remove variations caused by the strategy of choosing N spaces for training at random, we repeated the procedure for 10 rounds by choosing N different restrooms each round. Finally, we averaged the performances from the 10 rounds for each N training set size. The final results are shown in Figure 5.

As the number of training set size increases, the performance became more stable and gradually converged. The weighted F-Measure, which incorporates the precision and recall of both the restroom and non-restroom classes, converged at ~0.93. The weighted F-Measure fluctuated between 0.91 and 0.93 while the model was trained on less than 40 restrooms. This suggests that the model had not seen enough variations of restrooms at that point yet.

Generalizability of the Approach across Phones

Different phones may use different hardware. Furthermore, the microphone and the speaker can be placed at different distances and in different positions from one another. Thus, we must also validate that our approach generalizes to work on different phones. In addition to the Galaxy Nexus, we also collected additional restroom and non-restroom samples on a Nexus 4, an HTC One, and a Galaxy S. The data collected on these devices were not as extensive as the one collected on the Galaxy Nexus, but they helped to confirm the results obtained from analyzing the Galaxy Nexus data. Table 1 summarizes the amount of data collected on the four phones.

We performed a 10-fold cross validation using SVM as the classifier for each phone's data separately. The results are

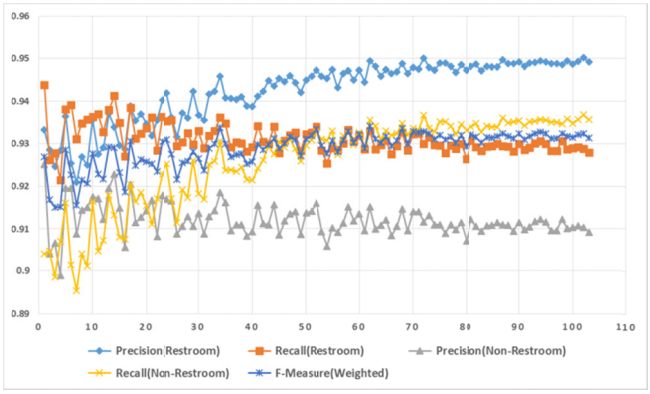


Figure 5. Measurements of the model's convergence with different number of restrooms used for training

Table 1. Sample data collected on the four phones

Phone Name	# of restrooms	#of restroom data samples	# of non-restroom data samples
Galaxy Nexus (GN)	103	4258	3216
Nexus 4	52	2230	1296
HTC One	20	600	523
Galaxy S	20	600	573

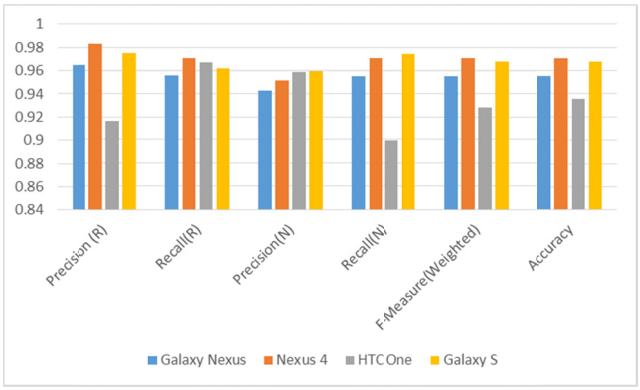


Figure 6. Model Generalization across phones: 10-fold Cross-validation Results on four phones separately (R: Restroom; N: non-restroom)

shown in Figure 6. It highlights that the classification models for the four different phones perform similarly well at between 0.92 and 0.98 (weighted F-Measure).

Unfortunately, different hardware settings along with software optimizations for the microphone and the speaker on different phones mean that the impulse responses captured by each phone are drastically different from each other. We applied the model trained on the Galaxy Nexus dataset and tested on the other three phones' dataset. The 10-fold cross validation results, shown in Figure 7, reveal that the extracted MFCCs features do not generalize across phones. Phone-independent classification does not perform as well as phone-dependent classification (weighted F-Measure between 0.43 and 0.63).

Effect of Occupancy & Sounds on the Model

We tested the model's robustness against the occupancy of a restroom and sound generated by the occupants in a restroom (*e.g.* urinating, flushing, and hand washing). Using the Galaxy Nexus, we collected new data from a restroom

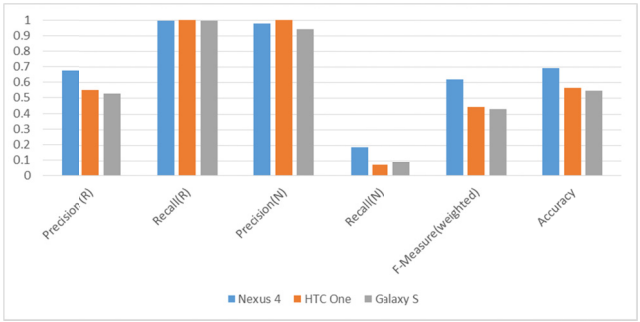


Figure 7. Model trained on Galaxy Nexus dataset and tested on the other three phones' dataset (R: restroom; N: non-restroom).

with two urinals, two toilet stalls, and three basins (seven functional spots in total) for the evaluation. We defined the occupancy rate as the percentage of functional spots occupied by people. Therefore, we had 14%, 29%, 43%, 57%, 71%, 86% and 100% occupancy rates. We collected 30 additional IR samples using the same method described earlier in the paper for each of the seven occupancy rates. In all seven cases, people simulated using the restroom in normal manners. We used the SVM classifier trained on the large Galaxy Nexus data corpus (described in the first row of Table 1) to test the new collected data. For the 14%, 29% and 86% occupancy rates, the model misclassified one sample each (accuracy: 97%), and correctly classified all samples for the remaining four occupancy rates.

Effect of Sweep Volume on the Model

We evaluated the influence of the sweep volume on the model to explore if the obtrusiveness of the sweep can be minimized by outputting the sweep at lower volumes. We collected 30 samples at 4 volume levels (0.1, 0.2, 0.5 and 0.75 of the max volume) from 13 additional restrooms using the Galaxy Nexus.

For comparison, we also extracted MFCCs using only the environment sound without outputting a sweep; this acted as the IR for the sweep at volume 0. We used the portion of the recording between the start of the recording and the start of the sweep (see Figure 2) as each space's environment sound. We note that a limitation and potential threat to validity in using this approach exists because the extracted data is ~0.15s in length each, in comparison to the 0.4 s recordings at other volumes. In this instance, because there is no sine wave sweep, we did not need to include an additional 0.3 s normally used to capture the IR after the sweep plays.

We performed a 10-fold cross validation on the data collected at each volume level. The results shown in Figure 8 demonstrates that models trained on recordings of IRs after an output sweep outputted at any of the four volume levels performed better than the model trained on recordings without a sweep; there was a 10% or more improvement in accuracy. Additionally, models trained on data captured when the application outputted sweeps at higher volumes generally performed better than those at lower volumes. Although models trained on IRs captured after lower volume sweeps lose some performances, they are still comparable in performance to those after higher volume sweeps. This suggested that it is possible to output the sweep at a lower volume in real practice so that the active probing is less perceivable and thus less obtrusive.

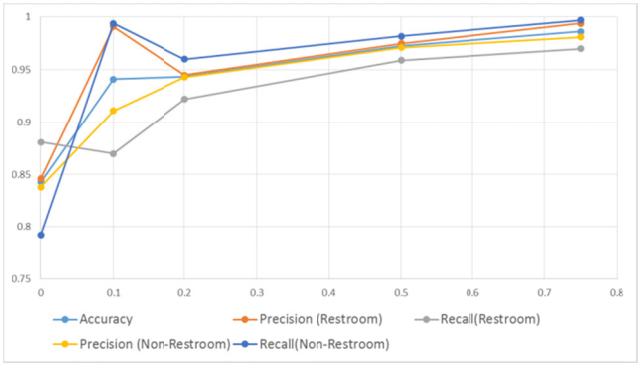


Figure 8. Measurements of the model using 5 different volume levels. Volume 0 uses only the portion without a sweep.

Table 2. Sample data collected in-the-wild using Galaxy Nexus.

Day	# of restrooms	#of restroom data samples	# of non-restroom data samples
1	9	378	788
2	4	171	485
3	3	141	390
4	4	109	413
5	11	704	2093

Continuous In-the-Wild Sampling and Evaluation

In this section, we performed continuous in-the-wild data collection to evaluate how well our classification approach works for realistic scenarios, such as when the user wears a Galaxy Nexus around the neck (see Figure 1). The data collection application plays a 0.1 second sine wave sweep sound at 0.3 of the max volume. The application automatically repeated this procedure in 5 seconds interval.

We collected data for about 2 hours per day on 5 different days. We collected data in different places each day and as a result the restrooms were all different. This approach enabled us to test the model's ability to classify new spaces. The temporal continuity of each day's data allowed us to apply the temporal optimization to the SVM classifier's predictions later. We followed the procedure described in the Feature Extraction section and again used SVM as the classifier. Table 2 summarizes the collected data (4169 non-restroom data and 1503 restroom data).

Classifying Spaces

We evaluated the model's ability to classify new restroom spaces in-the-wild. Because we collected data over 5 different days, the model could be trained on 1, 2, 3, 4, or 5 days' worth of data. For each number of days, we evaluated all the possible combinations of different days that could be grouped together. We had 5, 10, 10, 5 and 1 possible combinations for 1, 2, 3, 4 and 5 days' worth of data respectively. We used a 10-fold cross validation method to

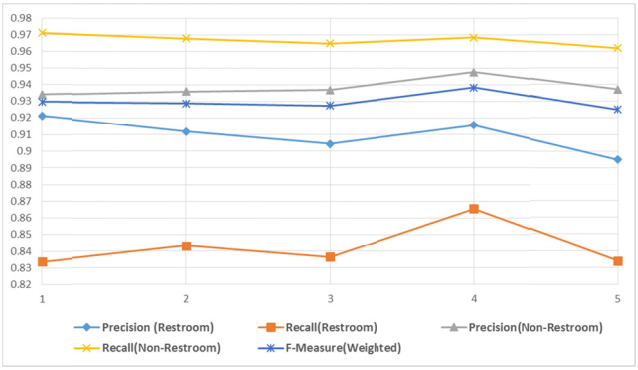


Figure 9. Measurements of the model trained on different number of days' data.

evaluate each combination. This tested the model's ability to accurately classify previously seen spaces as well as new ones, taking into consideration the fact that a person would likely to revisit some places and restrooms. Then, we averaged the 10-fold cross validation results of all combinations for a given number of days' worth of data to get the mean average values. The results, shown in Figure 9, illustrate that our model can predict new spaces with balanced precision and recall (weighted F-Measure > 0.92).

Improving Prediction Errors

In this section, we describe the prediction errors that affect the model's performance in classifying the continuous in-the-wild data that we collected and discuss how to correct them. The first type of error is a "spark," which is an isolated prediction of one class instead of the other. The second type of error is a "boundary" error, which happens when the user enters or leaves the restroom, but the prediction does not reflect that transition immediately. The third type of error is when the SVM predictions are "sporadic" and multiple wrong results are returned over a period of time.

Spark and sporadic prediction errors potentially can be eliminated to some extent by a smoothing algorithm, because it can be assumed that people normally do not jump in or out of the restroom for only 5 seconds (the sampling interval). However, boundary errors cannot be addressed by a smoothing algorithm.

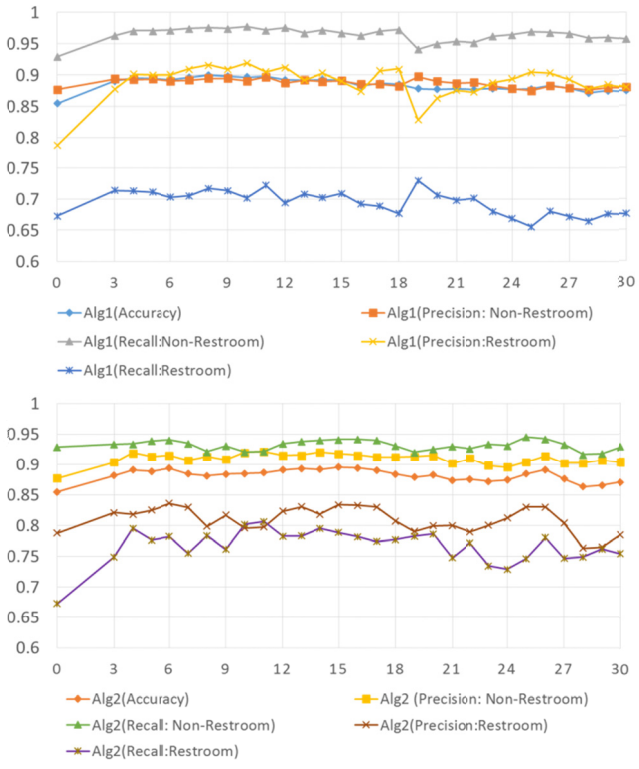


Figure 10. Evaluation results of the two smoothing algorithms with different window size (top: alg. 1; bottom: alg. 2)

Smoothing Algorithm I. The first algorithm keeps track of the current space type predicted by the SVM model. When a different space type is predicted by the model, then the algorithm will hold this new prediction and keep receiving further predictions from the model for a *pre-set window size* buffer *N*. If the majority of the predictions in the buffer match this different space type, then our algorithm will correct all buffered predictions as the new space type and also change the current space type that the user is in to this new one. If the majority of the buffered predictions still match the current space type, then all the buffered predictions will be classified as the current space type and the current space type remains unchanged.

Smoothing Algorithm II. Smoothing algorithm II is almost the same as the first one, except that it treats the transitions from restroom (non-restroom) to non-restroom (restroom) differently. People typically spend a small amount of their time in a restroom (people have reported spending approximately only 5 minutes in public restrooms [12]). Thus, to minimize potential misclassifications of actual transitions into a restroom space, algorithm *II* reduces the number of restroom predictions (> 1/3 * N instead of > ½ * N) that must be in the buffer window to correct an error.

We tested the two smoothing algorithms using window sizes from 3 to 30. Smoothing algorithms require the temporal continuity in the test data. Therefore, we used a leave one day's data out strategy for evaluation. For each one of the five combinations, we first trained the SVM model using four days' worth of data, and tested on the remaining day's data. Then, we applied our smoothing algorithms to the SVM prediction results. Finally we averaged the evaluation results of the five combinations for each window size. The performance results are reported in Figure 10. Window size zero means no smoothing algorithm was used. The difference in the performance compared to the one shown in Figure 9 is due to the fact that the cross-validation strategy allows the classifier to "see" a portion of each day's data during training phase while "leaving one day's data out" does not. We expect that performance will increase when trained with more days to increase the variations of restrooms. As the window size increases, the overall performance improves at first. Large window sizes, however, hurt the performance. One possible reason is that larger window sizes might cause more boundary errors during room type transition due to the majority voting strategy used in smoothing algorithms. Compared to the algorithm *I*, algorithm *II* improves the recall of restroom but sacrificed the precision of restroom.

CONCLUSION

In this paper, we described an infrastructure-independent method of detecting restrooms by actively probing the acoustics of an environment with the built in speaker and microphone on mobile phones. Our evaluations demonstrate that IRs captured after a sine wave sweep is outputted improve the accuracy of prediction compared

against only using the environment sound without a sweep. Models can be developed on different phones to classify new restrooms with a weighted F-Measure of 0.92~0.98. Occupancy, the presence of sounds, and the volume levels of the sweep do not affect the model's performance in significant ways. We discuss three types of errors that affect the prediction model and propose temporal smoothing algorithms to improve the prediction accuracy.

REFERENCES

1. Agnihotri, S., Rovet, J., Cameron, D., Rasmussen, C., Ryan, J. and Keightley, M. SenseCam as an everyday memory rehabilitation tool for youth with fetal alcohol spectrum disorder. In *Proc. SenseCam 2013,* ACM (2013), 86-87.
2. Beritelli, F. and Grasso, R. A Pattern Recognition System for Environmental Sound Classification based on MFCCs and Neural Networks. In *Proc. ICSPCS 2008,* IEEE (2008), 1-4.
3. Byrne, D., Doherty, A., Jones, G., F., Smeaton, A. Kumpulainen, S. and Järvelin. K. 2008. The SenseCam as a tool for task observation. In *Proc. BCS-HCI '08*, Vol. 2, 2008, 19-22.
4. Caprani, N., O'Connor, N., Gurrin, C. Experiencing SenseCam: a case study interview exploring seven years living with a wearable camera. In *Proc. SenseCam 2013*, ACM (2013), 52-59.
5. Chang, C., Lin, C. LibSvm: A library for support vector machines. *ACM Transactions on Intelligent Systems and Technology*, vol. 2(3), 2011, 188-205.
6. Farina, A. Simultaneous Measurement of Impulse Response and Distortion with a Swept-Sine Technique. *108th Convention of the Audio Engineering Society*, Paris, France. February 2000, 19-22.
7. Hempstalk, K., Frank, E., Witten, I. One-Class Classification by Combining Density and Class Probability Estimation. In *Proc. ECML PKDD 2008*, 505-519.
8. Hodges, S., Williams, L., Berry, E., Izadi, S., Srinivasan, J., Butler, A., Smyth, G., Kapur, N., Wood, K. SenseCam: A Retrospective Memory Aid. In *Proc. Ubicomp 2006,* ACM (2006), 177-193.
9. Keadle, S., Lyden, J., Hickey, A., Ray, E., Fowke, J., Freedson, P., Mattews, C. Validation of a previous day recall for measuring the location and purpose of active and sedentary behaviors compared to direct observation. *Int J Behav Nutr Phys Act*, 2014, 11:12.
10. Kerr, J., Marshall, S., Godbole, S., Chen, J., Legge, A., Doherty, A., Kelly, P., Oliver, M., Badland, H., Foster, C. Using the SenseCam to improve classifications of sedentary behavior in free-living settings. *American Journal of Preventive Medicine,* 44(3), 2013, 290–296.
11. Kelly, P., Marshall, S., Badland, H., Kerr, J., Oliver, M., Doherty, A., Foster, C. An ethical framework for automated, wearable cameras in health behavior research. *American Journal of Preventive Medicine*, 44(3), 2013, 314-319.
12. Kientz, J.A., Choe, E.K., Truong, K.N., Texting from the Toilet: Mobile Computing Use and Acceptance in Private and Public Restrooms. *Knowledge Media Design Institute, University of Toronto. Technical Report KMD-13-1.* April 2013.
13. Kunze, K., Lukowicz, P. Symbolic object localization through active sampling of acceleration and sound signatures. In *Proc. Ubicomp'07*, ACM (2007),163-180.
14. Kuttruff, H. Room acoustics. *CRC Press*, 2009, 160-292.
15. Marcu, G., Dey, A., Kiesler, S. Parent-driven use of wearable cameras for autism support: a field study with families. In *Proc. Ubicomp'12*. ACM (2012), 401-410.
16. Peltomen, V., Tuomi, J., Klapuri, A., Huopaniemi, J. Sorsa, T. Computational auditory scene recognition. In *Proc. ICASSP 2001*, IEEE (2001), 1941-1944.
17. Perina, A., Jojic, N. In the sight of my wearable camera: Classifying my visual experience. *Eprint arXiv: 1304. 7236*, 04/2013.
18. Piccardi, L.; Noris, B.; Barbey, O.; Billard, A.; Schiavone, G.; Keller, F.; von Hofsten, C., "WearCam: A head mounted wireless camera for monitoring gaze attention and for the diagnosis of developmental disorders in young children,". In *Proc. RO-MAN 2007,* IEEE (2007), 594-598.
19. Pirsiavash, H. and Ramanan, D. Detecting activities of daily living in first-person camera views. In *Proc. CVPR 2012*, IEEE (2012), 2847-2854.
20. Quattoni, A. and Torralba, A. Recognizing indoor scenes. In *Proc. CVPR'09*, IEEE (2009), 413–420.
21. Rossi, M.; Feese, S.; Amft, O.; Braune, N.; Martis, S.; Tröster, G. AmbientSense: A real-time ambient sound recognition system for smartphones. In *Proc. PerCom Workshop*, IEEE (2013), 230-235.
22. Rossi, M., Seiter, J., Amft, O., Buchmeier, S. and Tröster, G. RoomSense: an indoor positioning system for smartphones using active sound probing. In *Proc. AH 2013,* ACM (2013), 89-95.
23. Rossing, T., Moore, R., Wheeler, P. The Science of Sound, 3rd Edition. *Addison-Wesley*. 2001.
24. Stan, G., Embrechts, J., and Archambeau, D. Comparison of different impulse response meansurement techniques. *Journal of the Audio Engineering Society*, 50(4), 2002, 249-262.
25. Templeman, R., Korayem, M., Crandall, D., Kapadia, A. PlaceAvoider: Steering First-Person Cameras away from Sensitive Spaces. In *Proc. NDSS* 2014.
26. Tian, Y., Yang, X., Yi, C. and Arditi, A. Toward a computer vision-based wayfinding aid for blind persons to access unfamiliar indoor environments. *Machine Vision and Applications*, 24(3), 2013, 521-535.