Demo Abstract: Infrastructure-Free Smartphone Indoor Localization Using Room Acoustic Responses

Dongfang Guo[1,*], Wenjie Luo[1,*], Chaojie Gu[2,†], Yuting Wu[1], Qun Song[3], Zhenyu Yan[4,†], Rui Tan[1]
[1]Singtel Cognitive and Artificial Intelligence Lab for Enterprises, Nanyang Technological University
[2]College of Control Science and Engineering, Zhejiang University
[3]ERI@N, IGS, School of Computer Science and Engineering, Nanyang Technological University
[4]Department of Information Engineering, The Chinese University of Hong Kong

ABSTRACT

Smartphone indoor location awareness is increasingly demanded by a variety of mobile applications. The existing solutions for accurate smartphone indoor localization rely on additional devices or pre-installed infrastructure (e.g., dense WiFi access points, Bluetooth beacons). In this demo, we present EchoLoc, an infrastructure-free smartphone indoor localization system using room acoustic response to a chirp emitted by the phone. EchoLoc consists of a mobile client for echo data collection and a cloud server hosting a deep neural network for location inference. EchoLoc achieves 95% accuracy in recognizing 101 locations in a large public indoor space and a median localization error of 0.5 m in a typical lab area. Demo video is available at https://youtu.be/5si0Cq6LzT4.

CCS CONCEPTS

• **Human-centered computing → Smartphones**; • **Computing methodologies → Supervised learning by classification**.

ACM Reference Format:
Dongfang Guo[1,*], Wenjie Luo[1,*], Chaojie Gu[2,†], Yuting Wu[1], Qun Song[3], Zhenyu Yan[4,†], Rui Tan[1]. 2021. Demo Abstract: Infrastructure-Free Smartphone Indoor Localization Using Room Acoustic Responses. In *The 19th ACM Conference on Embedded Networked Sensor Systems (SenSys '21), November 15–17, 2021, Coimbra, Portugal.* ACM, New York, NY, USA, 2 pages. https://doi.org/10.1145/3485730.3492877

1 INTRODUCTION

Smartphone-based localization has attracted wide research in recent years due to the proliferation of smartphones and the increasingly demanded location-aware services. In outdoor environments, the global positioning system (GPS) can provide meter-accurate localization. However, GPS service are often unavailable in indoor environments due to the blockage of GPS signal and more stringent requirements of indoor positioning services.

∗ Co-primary authors listed in alphabetic order.
† This work was done when authors were at Singtel Cognitive and Artificial Intelligence Lab for Enterprises, Nanyang Technological University.

SenSys '21, November 15–17, 2021, Coimbra, Portugal

ACM ISBN 978-1-4503-9097-2/21/11…$15.00
https://doi.org/10.1145/3485730.3492877

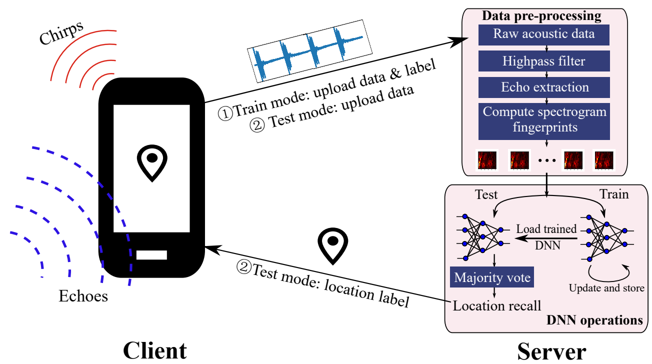


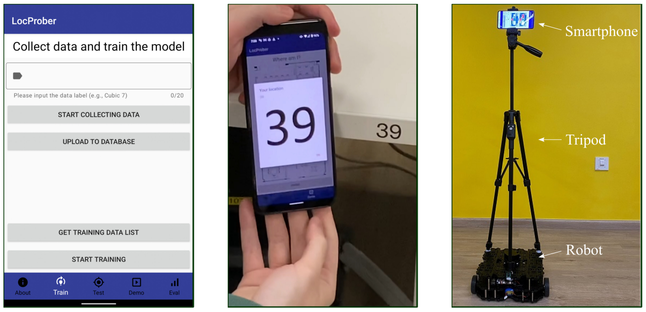
Figure 1: Workflow of EchoLoc.

Recent works exploit smartphone acoustic system for indoor localization [1–4]. However, [2, 3] only achieve room-level localization. [1, 4] only evaluate the localization performance at a few locations per room. We thereby demonstrate *EchoLoc*, a deep neural network (DNN) based system for accurate and large-scale smartphone indoor localization using room acoustic responses. EchoLoc comprises a mobile client and a cloud server. The mobile client controls the smartphone audio system to emit nearly inaudible chirps and record echoes. The cloud server hosting a DNN makes location inference and sends the result to the client. In two evaluated indoor environments, EchoLoc achieves a median localization error of 0.5 m on 61 locations in a 448 m^2 lab space and 95% location recognition accuracy on 101 locations in a large public space.

EchoLoc has four salient advantages. First, EchoLoc is infrastructure free and only requires the access to the smartphone audio system. We show that the potential applications of EchoLoc span from small meeting rooms to large semi-open concert halls. Second, EchoLoc achieves 95% location recall accuracy in a large public indoor space and sub-meter median localization error in the tested lab office. This localization accuracy can support a range of location-based mobile services such as navigation, way finding, workspace clock-in, artwork interaction, emergency response, etc. Third, EchoLoc works with normal hand grasping of the smartphone and require no special holdings of the phone. Lastly, the emitted chirp uses frequencies of 15 kHz to 20 kHz, which lies in the nearly inaudible range of human ears. Therefore, EchoLoc is robust to ambient noise and causes little/no annoyance to users.

2 SYSTEM OVERVIEW

Fig. 1 illustrates the workflow of EchoLoc.



(a) Developed mobile client (b) A successful location recognition (c) Customized data collection robot

Figure 2: EchoLoc client on the Android platform and a localization result in an office cubicle.

2.1 EchoLoc Client

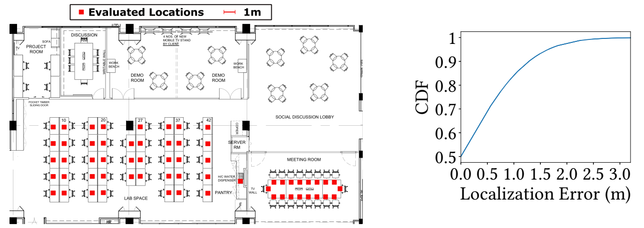
We developped a mobile application for training data collection and real-time localization. The client interface is shown in Fig. 2a. In the training mode, the client transmits chirps and records the corresponding acoustic traces of the fingerprinted positions, and then transfers the data and location label back to the server for model training. In addition, we build a robot-based data collector to facilitate the training data collection. As shown in Fig. 2c, a robot mounted with a phone client can navigate inside a room to collect the training data. In the inference mode, the client transmits a few chirps and transfers the collected acoustic traces to the server for location query. The predicted location from server is promptly showed on the smartphone screen. Fig. 2b demonstrates a correctly predicted location, in which we display the location number for easy visualization.

2.2 EchoLoc Server

We train and run the DNN at the cloud server following the pipeline as shown in the right side of Fig. 1. The DNN architecture used is ResNet-18, we customize the model so that it can accept acoustic spectrogram features as input. The server performs model training upon the client's request in the training mode, or performs the inference and sends the predicted location to the client in the inference mode. After obtaining the data from the client, the server first pre-processes the echo traces to generate echo features. The signal pre-processing includes filtering for inaudible band, echo traces extraction and echo spectrogram feature computation. If the server receives the training request from the client, or all the data for locations of interest have been collected, it will train the DNN model with the pre-processed fingerprinting data. Or, if the server is requested by the client to perform localization, it feeds the pre-processed fingerprints into the DNN, and outputs the corresponding prediction results. To achieve more reliable localization results, the server performs a majority vote based on the predicted results for multiple chirps at certain location, then returns the voting result to the client.

3 DEMONSTRATION

We implement EchoLoc client on the Android platform and deploy it on a Google Pixel 4 smartphone. The EchoLoc server is equipped with an NVIDIA Quadro RTX 6000 24GB GPU, with a DNN model



(a) Floor plan (b) Localization error CDF

Figure 3: Lab space and evaluation result.

built with PyTorch. The server system is based on Flask, a Python-based micro web framework. We use the Flask-RESTful extension to develop a set of representational state transfer (RESTful) application programming interfaces over Hypertext Transfer Protocol Secure (HTTPS) for client-server communications.

EchoLoc performance is evaluated in a public indoor space and a typical lab area.

■ **Public indoor space:** We deploy EchoLoc in a public indoor space. A total of 101 locations in this space are fingerprinted. Among the 101 locations, 50 locations are within rooms and the rest 51 are in corridors and foyers. For training, 800 fingerprints are collected using the smartphone at each location. Training data is sent to remote server for DNN model training. Upon completion, we evaluate the system performance by revisiting the locations where training data are collected. For evaluation, 100 test samples are collected at each location. EchoLoc achieves an overall test accuracy of 86%. Furthermore, we adopt the majority vote to improve the accuracy of our system. With 10 chirps to vote for the predicted location, the test accuracy increases to 95%. Noted that the data collection process with 10 chirps takes one second only.

■ **Lab area:** We evaluate EchoLoc in a $16 \times 28\,\mathrm{m}^2$ lab space. Its floor plan is shown in Fig. 3a. Fingerprints are collected at 61 locations within the lab, including 43 locations covering all cubicle seats and 18 chairs that are adjacent to each other in a meeting room. The locations are highlighted as red squares in Fig. 3a. Fig. 3b shows the cumulative distribution function (CDF) of the localization errors in meters. EchoLoc achieves a median error of 0.5 m. The 90%tile localization error is 1 m.

4 ACKNOWLEDGEMENT

This research was conducted at Singtel Cognitive and Artificial Intelligence Lab for Enterprises (SCALE@NTU), which is a collaboration between Singapore Telecommunications Limited (Singtel) and Nanyang Technological University (NTU) that is funded by the Singapore Government through the Industry Alignment Fund - Industry Collaboration Projects Grant.

REFERENCES

[1] Mirco Rossi, Julia Seiter, Oliver Amft, Seraina Buchmeier, and Gerhard Tröster. 2013. RoomSense: an indoor positioning system for smartphones using active sound probing. In *Proc. AHs*. 89–95.
[2] Qun Song, Chaojie Gu, and Rui Tan. 2018. Deep room recognition using inaudible echos. *Proc. IMWUT* (2018).
[3] Stephen P Tarzia, Peter A Dinda, and et al. Dick. 2011. Indoor localization without infrastructure using the acoustic background spectrum. In *Proc. MobiSys*.
[4] Yu-Chih Tung and Kang G Shin. 2015. EchoTag: Accurate infrastructure-free indoor location tagging with smartphones. In *Proc. MobiCom*. 525–536.