

Batphone: Audio based room shape detection using a smartphone

ABSTRACT

This is abstract.

1. INTRODUCTION

“... You took my sonar concept and applied it to every phone in the city. With half the city feeding you sonar, you can image all of Gotham.” - Lucious Fox, The Dark Knight.

When in the movie *The Dark Knight*, *Lucious Fox* mentioned the above dialogue to *Batman*, *Batman* was then trying to locate *Joker* to save the city of *Gotham*. There, *Batman* used the smartphones of people to use them as audio based imaging devices (based on the principle of sonar). It helped him to get real-time information about different events happening around the city. Interestingly, the idea of sonar stems from the *echolocation* citeXXX which bats or dolphins use to move in the dark or to catch their prey. *Echolocation* is a biological mechanism used to detect and locate objects by emitting usually high-pitched sounds that reflect off the object and carefully *listening* (via ears or other sensory receptors) to the different echoes. In this paper, we ask that can we give this superpower to humans through a smartphone ? Can we really *hear* the visual characteristics of our surrounding ? Can we just *plug and play* the mechanism of sonar *i.e.*echo based location technique to get the information ?

In this paper, while answering the above questions, we attempt to see how much of this *grand vision* can be achieved with our current smartphone without any external component or modification. We envision to map the surrounding with a smartphone using emitted and reflected audio signals. As a first step in that direction, specifically, we use smartphone to infer *2D* contour or shape skeleton of a room. We build a system called *BatPhone*, which user employs to get the shape of the room by moving around the room. To achieve this, our *Batphone* system utilizes FMCW radar based technology combined with IMU sensor based dead-reckoning. It also exploits the geometric constraints of the user movement and the room shape to construct an accurate room shape profile. Furthermore, we use specific reflection profile of FMCW chirps to detect different points of interest

like corners or wall intersections, which are critical for accurate profile.

In general, there is a recent surge of interest in industry to get structural information about the surrounding of a user. The reason is to enrich the augmented reality (AR) or virtual reality (VR) applications citeXXX. If these applications get accurate *3D* or *2D* mapping of the surrounding, they can provide more realistic and engaging experience to users. People can play AR games which can seamlessly blend with the surrounding (*e.g.*a spacecraft on a living room floor or an alien insect on a wall). Even VR experience becomes far smoother if it makes user aware of the surrounding citeXXX. To achieve this, there are multi-camera and multi-depth sensor based solutions like Microsoft Hololens, Google’s Project Tango tablet, Oculus VR headset etc. Unfortunately, this solution are costly and requires user to buy a separate device. So, our solution holds promise to provide cheap, device-free technique to get *2D* mapping or depth information by utilizing built-in speaker and microphone of a smartphone. Furthermore, *Batphone* can help to create different interesting use-cases like navigation app for blind people, an app for creating *3D* picture using image and depth information, enriching information for AR/VR applications etc. (as illustrated in Fig. 1).

Recently, researchers have become interested to use acoustic signals to build different tracking, gesture detection, or activity recognition applications citeXXX. This immense popularity of audio based applications is due to its easy implementation or prototyping. It is also due to the relative low speed of propagation and low sampling rate of audio signal compared to other high-frequency wireless signals. In general, there is no need of any hardware modification in current infrastructure. Therefore, these applications are easy-to-scale and cheap to implement. There are two basic threads of work are coming out of the research community : device-free mechanism and device-based mechanism. In device-based mechanism, the audio source and the receiver are different and for the device-free technique, the source and the receiver are collocated. In device-based mechanism, audio source like speaker emits specific audio chirps and mic of the

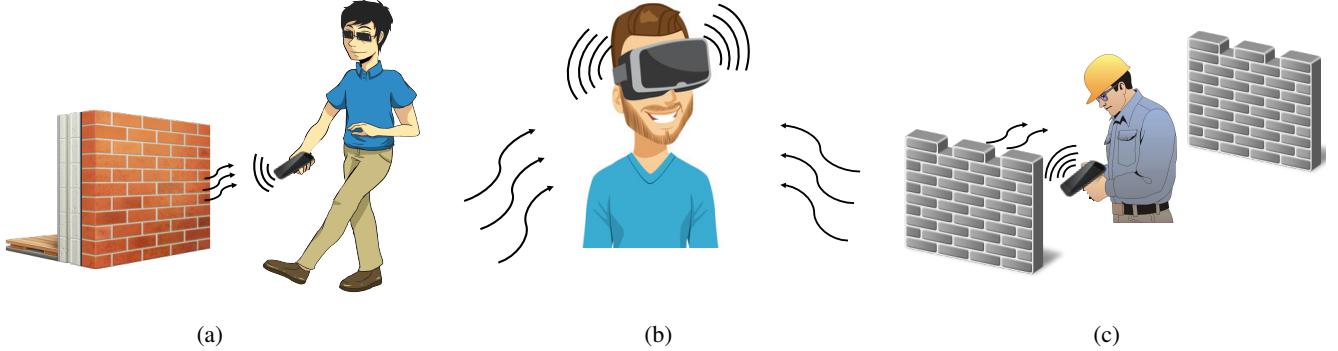


Figure 1: (a). Blind Person Obstacle Detection App. (b). AR/VR application. (c). Helping in construction mapping.

device like smartphone pick up that sound to analyze different properties like time-of-flight, phase citeXXX etc. Sometimes, the device collect these information from the emitted signal of multiple speakers at different locations to localize accurately citeXXX, tonetrack,DopEnc. In device-free mechanism, researchers use smartphone or smartwatch as transceiver and pick up the movements of finger or hand. Authors have also exploited doppler based frequency change, FMCW chirp based frequency change, or channel impulse response estimation based techniques citeXXX to track fine-grained movement in presence of varying degree of multi-path. In both of the schemes, unsurprisingly, smartphones are used as the central device to render the application.

To the best of our knowledge, we are first to propose a single step infrastructure-free smartphone based acoustic mapping solution which does not depend upon any crowd-sourcing or multiple trials by same user. In this paper, as a first step to build an acoustic based indoor mapping solution, we attempt to make a $2D$ contour of a closed or semi-closed space. To achieve this, firstly we analyze the capability of the mics and speakers current smartphone which may be critical in this application. We carefully investigate different trade-offs involved in designing a light-weight calibration-free solution which suits the purpose. Secondly, we borrow the principle of RADAR citeXXX technology based on Frequency Modulated Continuous Wave (FMCW) technology to calculate the distances from the walls to user. We also show why the echo-based method citeXXX or channel impulse response (CIR) based methods are difficult to use in this problem, specifically in a calibration-free single-step setting. We carefully look at the challenges involved in multi-path heavy indoor environment and reflection dominant (resulting in numerous echoes) audio signals, to achieve that goal. It is important to point out that, all the audio based tracking related works [1, 2, 3] cancel out the static multi-path by taking difference of consecutive samples. Whereas we exploit this multi-path which indirectly manifests itself in the FMCW profile to get critical structural information like corner or clutter. Furthermore, we employ customized IMU-based dead-reckoning

combined with these distance measurements and systematic geometric constraints to derive different surrounding shape information. We slowly build from a single wall scenario to multi-wall setting in the presence of static clutter. We also show that with other possible *easy-to-attach* extensions to the smartphone, we can make the system faster and more light-weight. Finally, we demonstrate the strength of our core technology through a set of applications.

Our main contributions are the following:

- We measured in detail the building blocks necessary to design a infrastructure-free, calibration-free, and device-free acoustic mapping solution. Specifically, we focus on exploring different parameters, and trade-offs of a FMCW-based distance measurement scheme, which is particularly suitable for mapping, on a smartphone .
- We developed a novel algorithm which can construct the contour of any indoor space around a user with a smartphone in a calibration-free manner.
- We also build an *end-to-end* system using our algorithm which intelligently combines IMU-based dead-reckoning and FMCW-based surrounding profiling, to construct the contour of an indoor space around a user.

Section Description.

2. FMCW PRIMER

Frequency Modulated Continuous Wave (FMCW) radar technology is a range and velocity measuring technique, in which transmitter transmits a frequency modulated chirped signal and the receiver uses the frequency shift between the transmit signal and the received signal to estimate the range.

In our setup we have co-located speaker and microphone. The speaker transmit an FMCW chirp periodically as shown in Figure XXX. The frequency of the signal sweeps linearly from f_{min} to f_{max} in each period. The frequency within each sweep is $f = f_{min} + \frac{Bt}{T}$, where B is the signal bandwidth, T is the sweep time. Integrating the frequency over

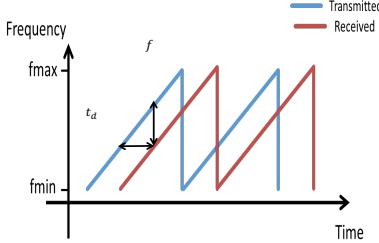


Figure 2: FMCW in action

time gives us phase of the signal: $u(t) = 2\pi(f_{min}t + B\frac{t^2}{2T})$. Thus giving the transmitted signal during the n-th sweep to be

$$v_t(t') = \cos(2\pi f_{min}t' + \pi B\frac{t'^2}{T})$$

where $t' = t - nT$.

Lets consider the transmitted chirp signal is reflected by the object and arrives at the receiver after a delay t_d . The received signal is attenuated and delayed in time, and becomes:

$$v_r(t) = \alpha \cos(2\pi f_{min}(t - t_d) + \frac{\pi B(t - t_d)^2}{T})$$

where α is the attenuation factor.

The receiver mixes (i.e., multiplies) the received signal with the transmitted signal. That is, $v_m(t) = v_r(t)v_t(t)$. Thus, $v_m(t)$ is a product of two cosines. By using $\cos A \cos B = (\cos(A - B) + \cos(A + B))/2$ and filtering out the high frequency component $\cos(A + B)$, $v_m(t)$ becomes:

$$v_m(t) = \alpha \cos(2\pi f_{min}t_d + \pi \frac{B(2t_d - t_d^2)}{T}).$$

Suppose the mobile is at distance $\frac{R}{2}$ from the reflecting object, reflection from that object would propagate a distance of R before it is received by the microphone, so t_d is given by $(R)/c$. Plugging t_d into signal above, gives us

$$v_m(t) = \alpha \cos(2\pi f_{min} \frac{R}{c} + 2\pi Bt \frac{R}{cT} - \pi B \frac{R^2}{c^2 T})$$

To analyze the frequency components of the above mixed signal we take derivative of its phase, the constant term can be ignored and the terms quadratic with respect to $1/c^2$ are too small and can also be ignored. The remaining frequency component, denoted as f_p comes out to be:

$$f_p = \frac{1}{2\pi} \frac{\delta \text{Phase}}{\delta t} = \frac{BR}{cT}$$

This means that the frequency spectrum of the mixed signal will have a peak at $\frac{BR}{cT}$. If there are multiple propagation paths between the transmitter and the receiver, multiple peaks are observed in the FFT spectrum of the mixed signal. Based on observed f_p in the FFT spectrum, the distance R

can be derived as:

$$R = \frac{f_p c T}{B}$$

In our set up, before being transmitted the FMCW chirp is shaped by passing it through a window. Shaping the chirp improves its signal-to-noise ratio by improving the peak-to-sidelobe ratio. The chirp signal is windowed using a hanning window XXX ref. The hanning window is described by

$$H[n] = 0.5 * \left(1 - \cos\left(\frac{2\pi n}{N-1}\right)\right). \quad (1)$$

The discretized pulse is shaped by performing an element-wise multiplication between the discretized pulse vector and the discretized hanning window. Finally, the discretized shaped pulse is sent to the speaker's buffer to be transmitted

3. BATPHONE : SYSTEM OVERVIEW

Fig. 3 illustrates how the system will work.

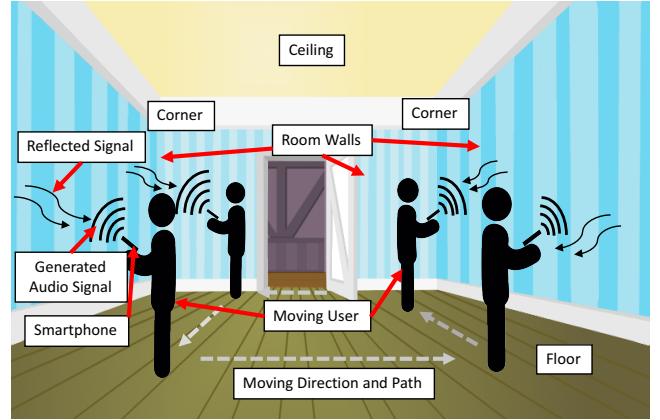


Figure 3: Batphone

Following are the building blocks coupled with their description :

3.0.1 Dead Reckoning

- Step size estimation
- Step detection and count
- Heading direction
- Phone orientation

3.0.2 FMCW based distance

- FMCW parameter choices (Chirp duration, Mic choice, Frequency response of speaker and mic based selection, Gap between chirps, Bandwidth selection, Sampling rate selection (Higher means lower quantization error or can also be done with zero padding).)

- Synchronization with dead reckoning.
- FMCW profile based corner detection and clutter filtering.

3.0.3 Room shape profiling

- Geometric constraints on the room wall and movement.
- Corner detection help in profiling.

4. DESIGN CHOICES

4.1 Different False Attempts or Starts

- Pulse based echo selection and finding or Echo-sorting (Too much direct path impact).
- Angle-of-arrival calculation (Too less number of mics and sine based or CIR based phase measurement won't work due to high level of multi-path).
- MUSIC based peak selection (Prior knowledge of number of peaks).

4.2 FMCW based observations

4.2.1 Resolution

It depends upon the bandwidth used.

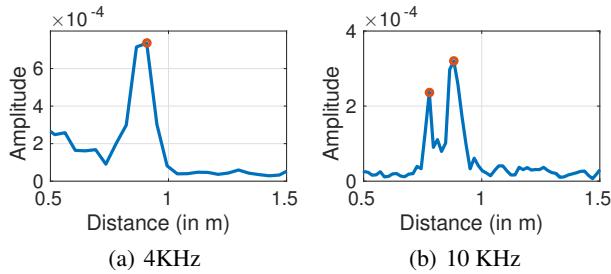


Figure 4: Impact of different bandwidths on resolution FMCW based distance measurement

4.2.2 Volume Level

It depends upon the proper volume level used.

4.2.3 FMCW distance measurement with sizes

It depends upon the reflection measurement with sizes.

- Junction causes multiple peaks.
- Maximum peak in FMCW corresponds to the main wall distance.
- Always get perpendicular distance from wall.
- Peak amplitude increases with the size of the reflector and it saturates after certain size.

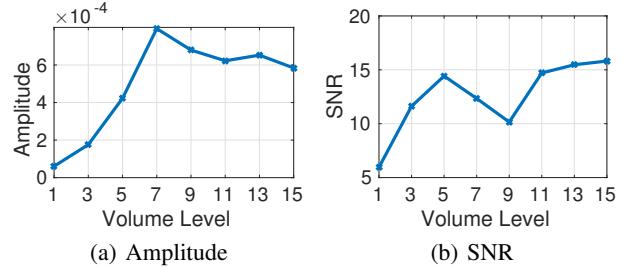


Figure 5: Impact of different volume level on FMCW based distance measurement

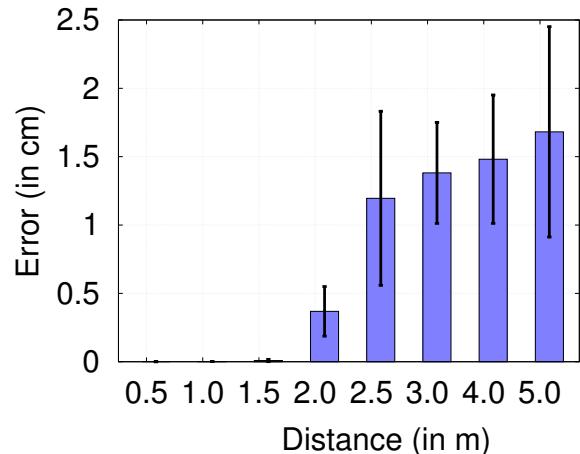


Figure 6: FMCW error change with distance

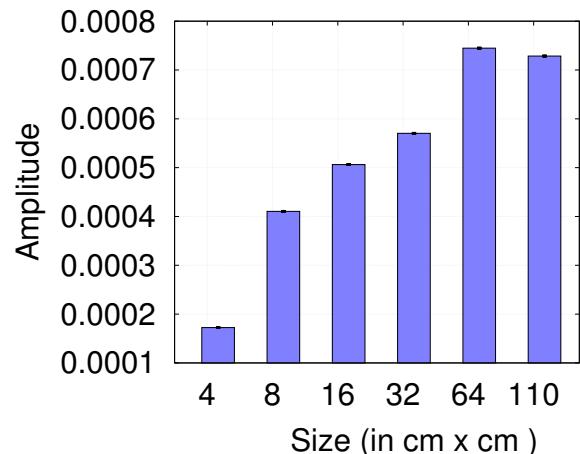


Figure 7: FMCW amplitude with different sizes

- Corner peaks are significantly different (at least for corner less than or equal to 90 degree). There are multiple peaks and peak spread is more.
- Audio lobe direction and mic position matters.

5. ALGORITHM DESCRIPTION

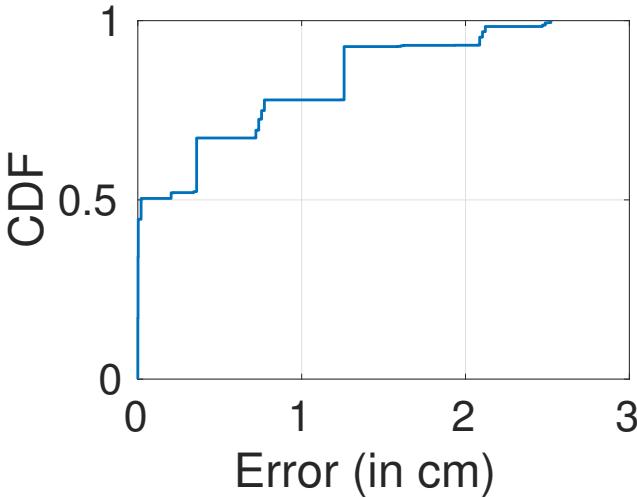


Figure 8: CDF of FMCW based distance calculation error.

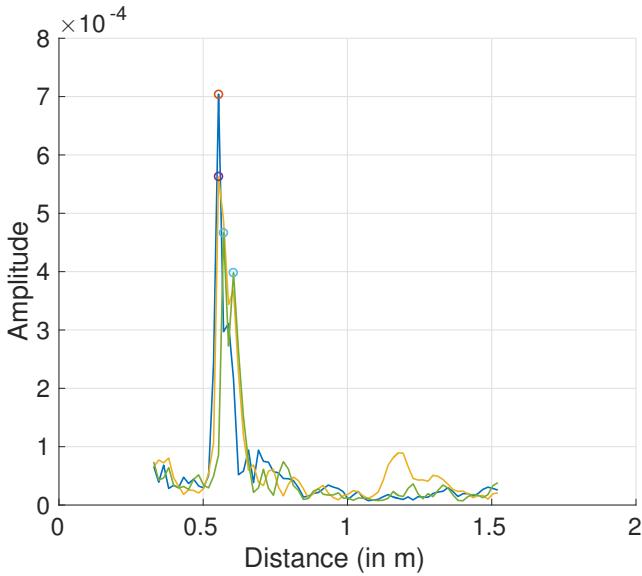


Figure 9: Perpendicular distance measured at different angles.

Description of the algorithm and geometric arguments with some explanatory figures. Fig. 10 illustrates the algorithm used in room shape detection.

6. GEOMETRIC CONSTRAINTS

Once we have synchronized position samples (x_t, y_t) from dead-reckoning and range readings (r_t) from FMCW, we now need to estimate the coordinates of the points on the wall that we received reflection $((x'_t, y'_t))$ from and those points we estimate the wall. Since the range values are perpendicular distances from the wall we get the following two geometric constraints on the values of (x'_t, y'_t) .

$$(y'_t - y_t)^2 + (x'_t - x_t)^2 = r_t^2$$

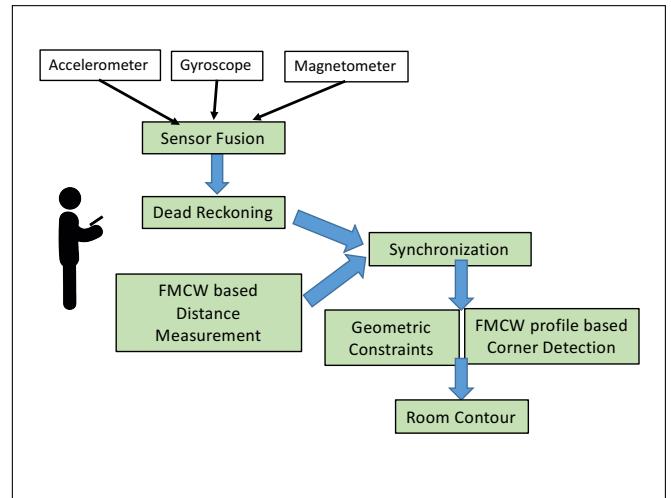


Figure 10: Algorithm sketch of *Batphone*

Algorithm 1 BATPHONE-SINGLE

```

1:  $x_0 \leftarrow$  current phone orientation
2:  $p(x) \leftarrow$  current position of the person
3: procedure STEP DETECTION( $\mathbf{a}$ )
4:   if  $Acc > Th$  then return True
5:   return False
6: procedure GETDEADRECKONING
7:    $t_0 \leftarrow$  CURRENTSYSTEMTIME
8:    $t \leftarrow t_0$ 
9:    $w \leftarrow$  EMPTYVECTOR
10:  return True
11: procedure UPDATERELOCATION( $x_0, w$ )
12:    $x_{new} \leftarrow \arg \min_{x_{min} < x < x_{max}} \text{MINDTW}(x_0, x, w)$ 
13:   return  $x_{new}$ 
  
```

$$\frac{y'_t - y_t}{x'_t - x_t} = \frac{-1}{m_2}$$

Where m_2 is the gradient of wall. To calculate m_2 we get the wall orientation with respect to the path traversed by the phone, as shown in fig XXX, is given as $\theta = \sin^{-1}(\frac{d_t}{X})$. Using similar triangle property $X = \frac{d_t \sqrt{(y_{t+1}-y_t)^2 + (x_{t+1}-x_t)^2}}{d_{t+1}-d_t}$, thus giving us

$$\theta = \sin^{-1}\left(\frac{d_{t+1}-d_t}{\sqrt{(y_{t+1}-y_t)^2 + (x_{t+1}-x_t)^2}}\right)$$

m_2 can now be calculated as $m_2 = \tan(\theta + \tan^{-1}(m_1))$, or $m_2 = \tan(-\theta + \tan^{-1}(m_1))$ depending on the side of the path the reflecting wall is present. Here $m_1 = \frac{y_{t+1}-y_t}{x_{t+1}-x_t}$ is the gradient of the path traversed.

Using these constraints x'_t and y'_t can be calculated as

$$x'_t = \pm \sqrt{r_t^2 * \frac{m_2^2}{m_2^2 + 1}} + x_t;$$

$$y'_t = (x_t - x'_t)/m_2 + y_t$$

This would give us four set of values for x'_t and y'_t , two on each side of the path. Using the phone orientation we know which side of the path the wall is enabling us to filter out two set of values. Among the remaining two we pick the set of values which would satisfy the following simple constraint

$$\frac{y'_{t+1} - y'_t}{x'_{t+1} - x'_t} = m_2$$

7. SETUP

In this section, we describe our different experimental setups in detail.



Figure 11: Contour Experiment



Figure 12: (a). Anechoic Experiment 1. (b). Anechoic Experiment 2.

8. METRICS

9. EVALUATION

This is result section.

9.1 Dead-reckoning based estimation

- Path profiling error (accelerometer + gyro)

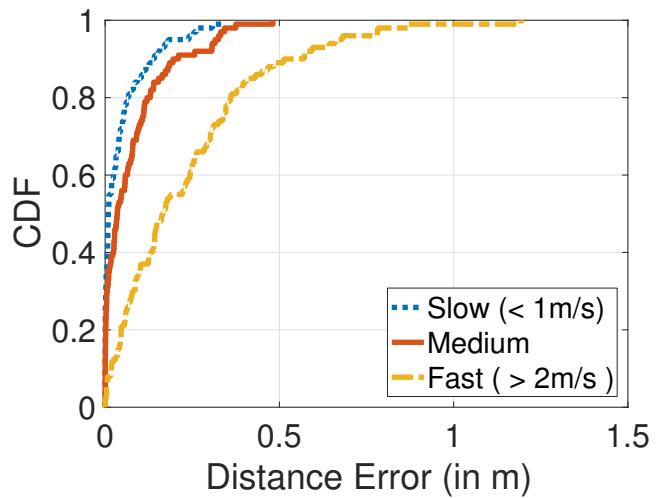


Figure 13: Dead-reckoning based distance estimation error.

9.2 Single wall Contour

- Different shape Straight and Angled
- Also need to try more shapes and sizes

9.3 Multi wall Contour

Corner detection accuracy (False positive rate)

Closed Intersections

- Different angles
- Different sized walls in intersection
- Direction association accuracy (If you are moving along a straight line and you have one wall in right or left side. How do you know which side is your wall ?)

Semi-Closed intersections

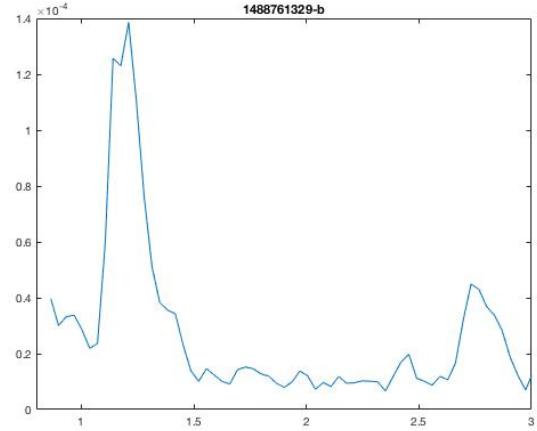


Figure 14: FMCW based and peak detection for distance measurement

9.4 Multi wall Contour with clutter

Sensitivity to surrounding small objects by trying to add

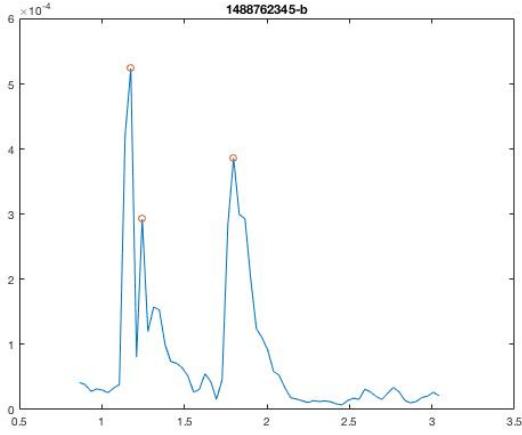


Figure 15: With floor

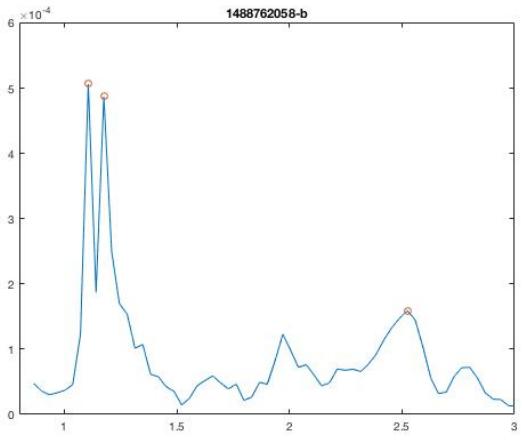


Figure 16: Without floor

different sizes small objects at different locations to see if you can effectively filter them out

10. POINTS OF DISCUSSION

Following are the points we learned.

11. RELATED WORK

Acoustic echoes reveal room shape [4, 5, 6, 7]

Finding source of audio [8]. Acoustic signals improve 3D visualization [9].

Sound-based smartphone proximity detection [10]. Sound based encounter detection [11]. Ambient sound finger-printing based room detection [12]. Sound-reflection signature based position detection [13].

Software sonar sensor [14], Floor-plan with ultra-sonic sensor [15],

Tracking (FingerIO [1], CAT [3], aamouse [2] [16]).

Ambient sound based fingerprinting. [17, 18] and sound

based indoor localization [19, 20].

Infrastructure-based acoustic indoor localization system [21, 22, 23, 24, 25, 26].

Smartphone-based 2D-mapping [14, 27].

Audio based phone range-finder [28].

12. CONCLUSION

This is Conclusion.

13. REFERENCES

- [1] R. Nandakumar, V. Iyer, D. Tan, and S. Gollakota, “Fingerio: Using active sonar for fine-grained finger tracking,” in *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI ’16, pp. 1515–1525, 2016.
- [2] S. Yun, Y.-C. Chen, and L. Qiu, “Turning a mobile device into a mouse in the air,” in *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*, MobiSys ’15, pp. 15–29, 2015.
- [3] W. Mao, J. He, and L. Qiu, “High-precision acoustic motion tracking,” in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, MobiCom ’16, pp. 491–492, 2016.
- [4] R. Scheibler, I. Dokmanić, and M. Vetterli, “Raking echoes in the time domain,” in *2015 IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 554–558, April 2015.
- [5] I. Dokmanić, R. Parhizkara, A. Walthera, Y. M. Lub, and M. Vetterli, “Acoustic echoes reveal room shape,” *Proceedings of National Academy of Sciences (PNAS)*, 2013.
- [6] M. Krekovic, I. Dokmanic, and M. Vetterli, “Echoslam: Simultaneous localization and mapping with acoustic echoes,” in *2016 IEEE International Conference on Acoustics, Speech and Signal Processing, ICASSP 2016, Shanghai, China, March 20-25, 2016*, pp. 11–15, IEEE, 2016.
- [7] M. Krekovic, I. Dokmanic, and M. Vetterli, “Omnidirectional bats, point-to-plane distances, and the price of uniqueness,” *CoRR*, vol. abs/1609.05512, 2016.
- [8] E. Kirby, S. Park, Y. Wang, and Y. Chen, “Hearhere: Smartphone based audio localization using time difference of arrival: Demo,” in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, MobiCom ’16, pp. 509–510, 2016.
- [9] W. Hussain, J. Civera, and L. Montano, “Grounding acoustic echoes in single view geometry estimation,” in *Proceedings of the Twenty-Eighth AAAI Conference on Artificial Intelligence*, AAAI’14, pp. 2760–2766, AAAI Press, 2014.
- [10] B. Thiel, K. Kloch, and P. Lukowicz, “Sound-based proximity detection with mobile phones,” in

- Proceedings of the Third International Workshop on Sensing Applications on Mobile Phones, PhoneSense '12*, (New York, NY, USA), pp. 4:1–4:4, ACM, 2012.
- [11] H. Zhang, W. Du, P. Zhou, M. Li, and P. Mohapatra, “Dopenc: Acoustic-based encounter profiling using smartphones,” in *Proceedings of the 22Nd Annual International Conference on Mobile Computing and Networking*, MobiCom ’16, pp. 294–307, 2016.
 - [12] S. P. Tarzia, P. A. Dinda, R. P. Dick, and G. Memik, “Indoor localization without infrastructure using the acoustic background spectrum,” in *Proceedings of the 9th International Conference on Mobile Systems, Applications, and Services*, MobiSys ’11, pp. 155–168, 2011.
 - [13] Y.-C. Tung and K. G. Shin, “Echotag: Accurate infrastructure-free indoor location tagging with smartphones,” in *Proceedings of the 21st Annual International Conference on Mobile Computing and Networking*, MobiCom ’15, pp. 525–536, 2015.
 - [14] G. Daniel, S. George, T. David, and Z. Gang, “A software based sonar ranging sensor for smart phones,” *IoT*, 2015.
 - [15] K. Yukitoshi, A. Yutaka, and Y. Keiichi, “A floor plan creation tool utilizing a smartphone with an ultrasonic sensor gadget,” in *13th IEEE Annual Consumer Communications & Networking Conference, CCNC 2016, Las Vegas, NV, USA, January 9-12, 2016*, pp. 131–136, 2016.
 - [16] W. Ruan, Q. Z. Sheng, L. Yang, T. Gu, P. Xu, and L. Shangguan, “Audiotest: Enabling fine-grained hand gesture detection by decoding echo signal,” in *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, UbiComp ’16, pp. 474–485, 2016.
 - [17] H. Lu, W. Pan, N. D. Lane, T. Choudhury, and A. T. Campbell, “Soundsense: Scalable sound sensing for people-centric applications on mobile phones,” in *Proceedings of the 7th International Conference on Mobile Systems, Applications, and Services*, MobiSys ’09, pp. 165–178, 2009.
 - [18] M. Azizyan, I. Constandache, and R. Roy Choudhury, “Surroundsense: Mobile phone localization via ambience fingerprinting,” in *Proceedings of the 15th Annual International Conference on Mobile Computing and Networking*, MobiCom ’09, pp. 261–272, 2009.
 - [19] J. Ruoxi, J. Ming, C. Zilong, and J. S. Costas, “Soundloc: Accurate room-level indoor localization using acoustic signatures,” in *IEEE International Conference on Automation Science and Engineering, CASE 2015, Gothenburg, Sweden, August 24–28, 2015*, pp. 186–193, 2015.
 - [20] M. Rossi, J. Seiter, O. Amft, S. Buchmeier, and G. Tröster, “Roomsense: an indoor positioning system for smartphones using active sound probing.,” in *AH*, pp. 89–95, ACM, 2013.
 - [21] F. Höflinger, R. Zhang, J. Hoppe, A. Bannoura, L. M. Reindl, J. Wendeberg, M. Bührer, and C. Schindelhauer, “Acoustic self-calibrating system for indoor smartphone tracking (ASSIST),” in *2012 International Conference on Indoor Positioning and Indoor Navigation, IPIN 2012, Sydney, Australia, November 13-15, 2012*, pp. 1–9, 2012.
 - [22] T. Janson, C. Schindelhauer, and J. Wendeberg, “Self-localization application for iphone using only ambient sound signals,” in *2010 International Conference on Indoor Positioning and Indoor Navigation*, pp. 1–10, Sept 2010.
 - [23] A. Mandal, C. V. Lopes, T. Givargis, and A, “Beep: 3D indoor positioning using audible sound,” *2005 Second IEEE*, 2005.
 - [24] C. Peng, G. Shen, Y. Zhang, Y. Li, and K. Tan, “Beepbeep: A high accuracy acoustic ranging system using cots mobile devices,” in *Proceedings of the 5th International Conference on Embedded Networked Sensor Systems*, SenSys ’07, (New York, NY, USA), pp. 1–14, ACM, 2007.
 - [25] I. Rishabh, D. Kimber, and J. Adcock, “Indoor localization using controlled ambient sounds,” in *2012 International Conference on Indoor Positioning and Indoor Navigation (IPIN)*, pp. 1–10, Nov 2012.
 - [26] G. Borriello, A. Liu, T. Offer, C. Palistrant, and R. Sharp, “Walrus: Wireless acoustic location with room-level resolution using ultrasound,” in *Proceedings of the 3rd International Conference on Mobile Systems, Applications, and Services*, MobiSys ’05, (New York, NY, USA), pp. 191–203, ACM, 2005.
 - [27] D. Graham, G. Zhou, E. Novak, and J. Buffkin, “A smartphone compatible sonar ranging attachment for 2-d mapping,” *IoT*, 2016.
 - [28] J. H. Gao and L.-S. Peh, “A smartphone-based laser distance sensor for outdoor environments,” in *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pp. 2922–2929, May 2016.