Learning Human Context through Unobtrusive Methods

Yanyong Zhang

WINLAB, Rutgers University

yyzhang@winlab.rutgers.edu

RUTGERS
WINLAB | Wireless Information Network Laboratory

We care about our contexts

Meeting Vigo: your first energy meter

Fitbit: Get Fit, Sleep Better, All in the one

Fall detection for the elderly

But,

Can we learn contexts in an unobtrusive manner?

- No need to wear a device
- No need to report status
- No extensive calibration
- It naturally takes place as we live our life

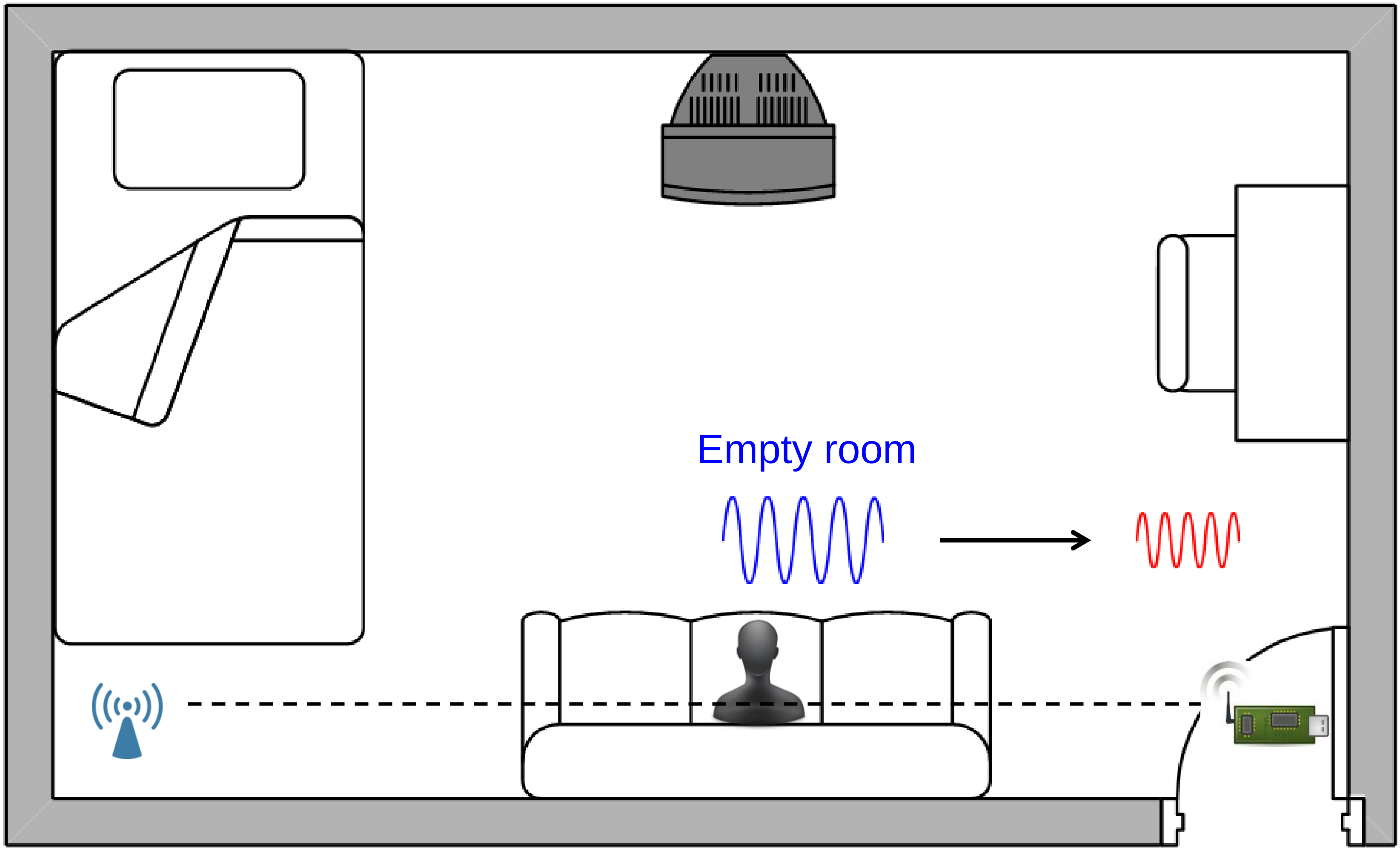
SCPL

Radio-frequency (RF) based device-free localization: location, trajectory, speed

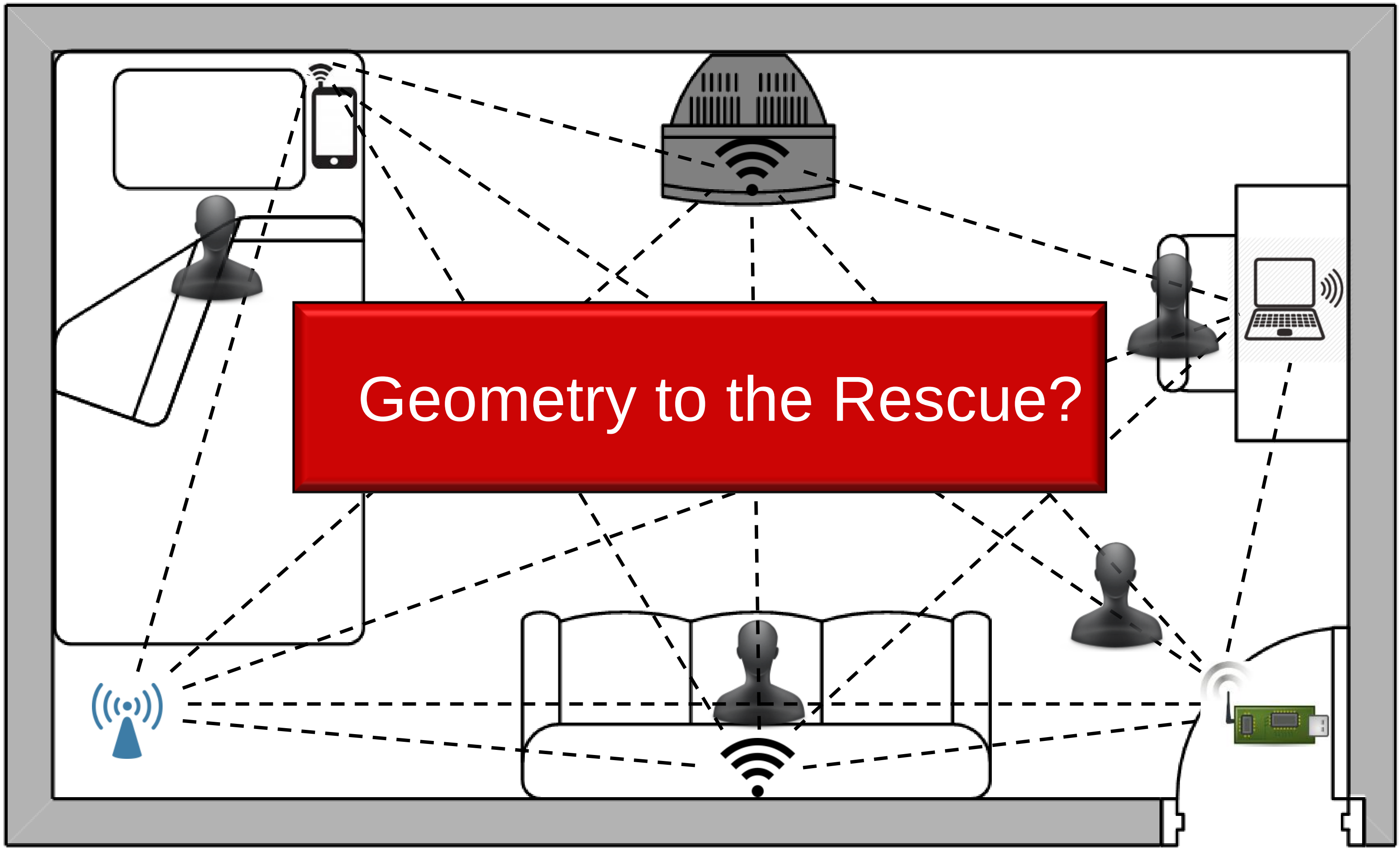
[1] C. Xu, B. Firner, Y. Zhang, R. Howard, J. Li, and X. Lin. Improving rf-based device-free passive localization in cluttered indoor environments through probabilistic classification methods, In ACM/IEEE IPSN, 2012.

[2] C. Xu, B. Firner, R.S. Moore, Y. Zhang, W. Trappe, R. Howard, F. Zhang, and N. An. Scpl: indoor device-free multi-subject counting and localization using radio signal strength. In ACM/IEEE IPSN, 2013.

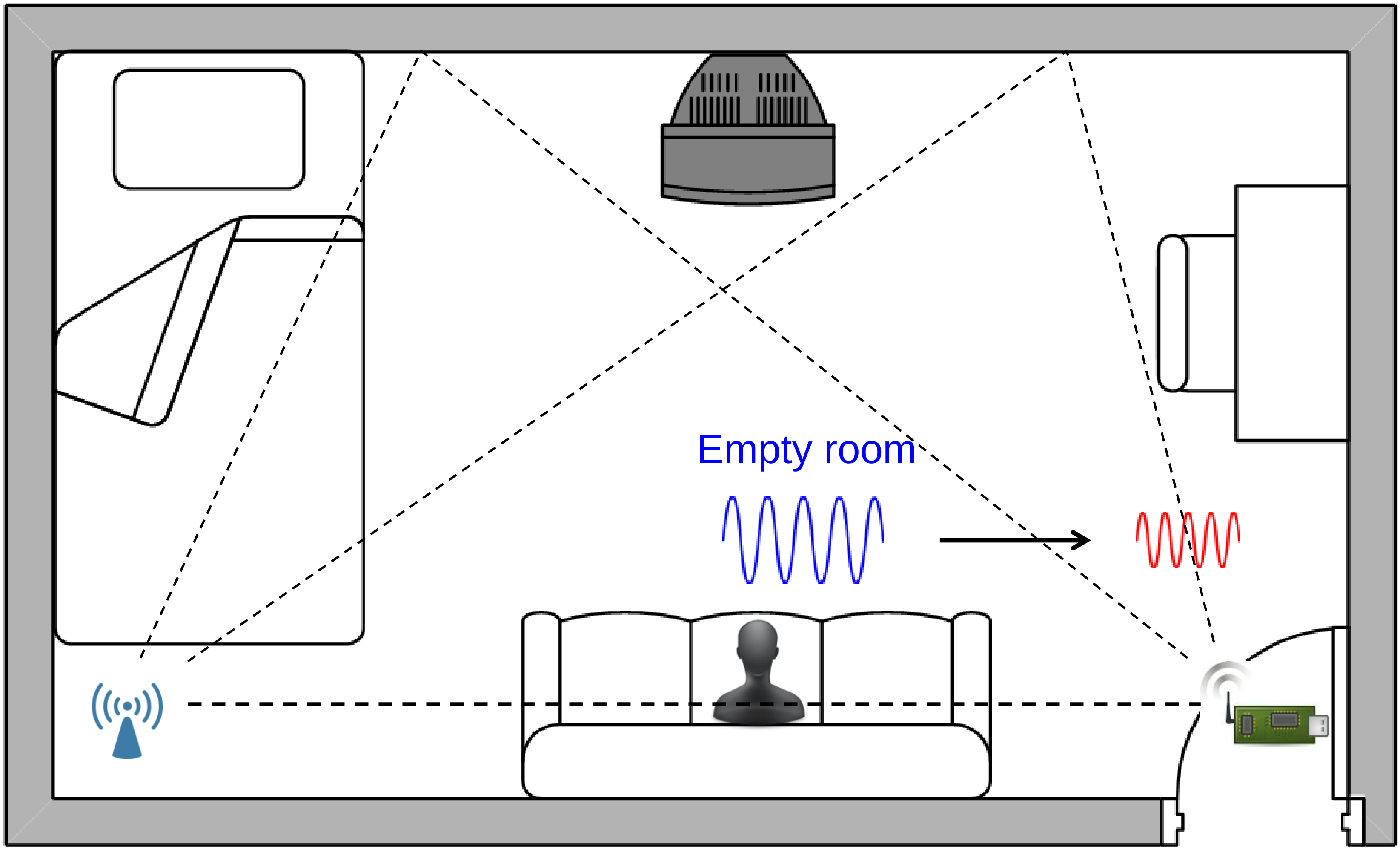
Device Free Passive Localization



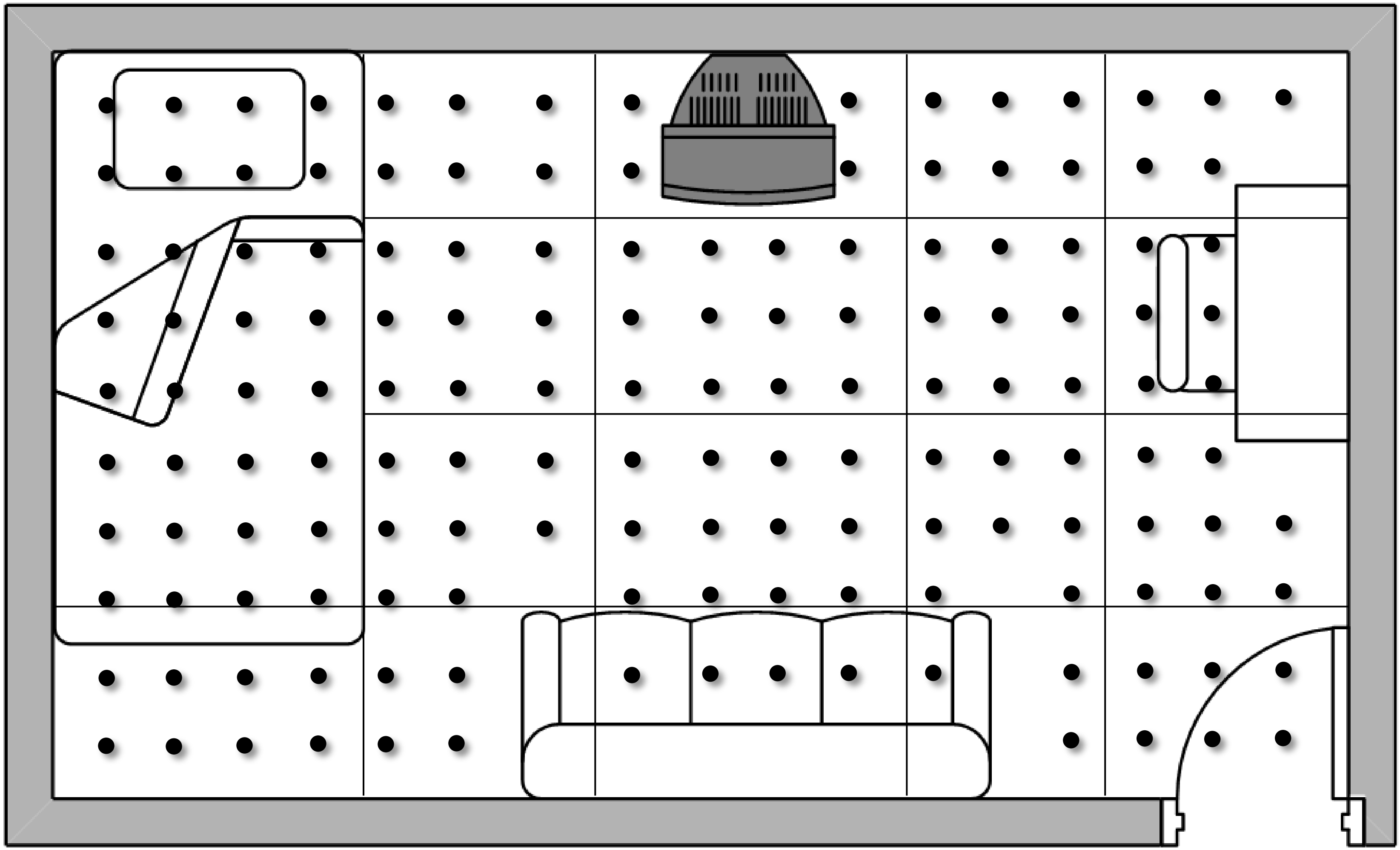
DfP Localization



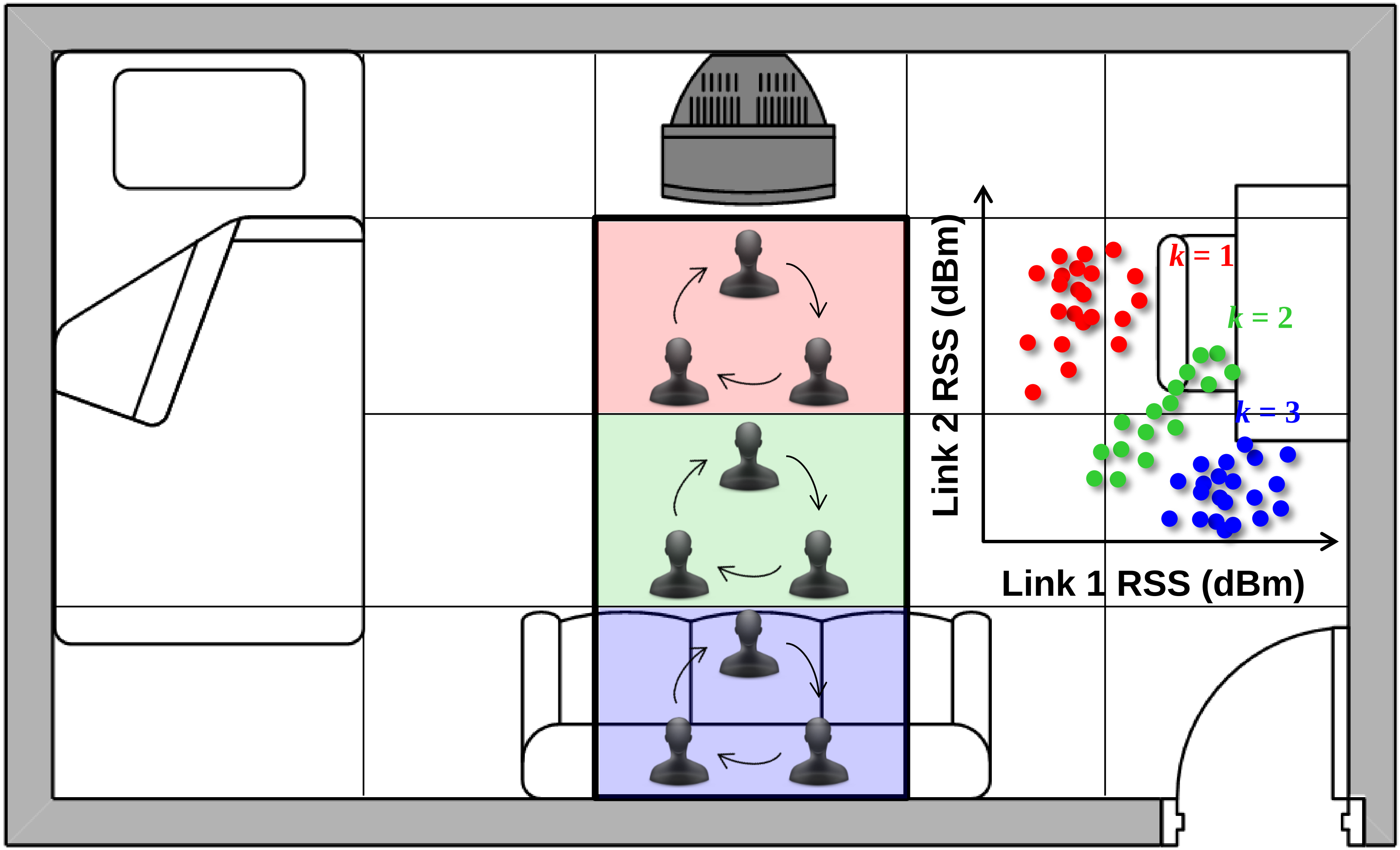
No! Because of Multi-path effect



Fingerprinting



Cell-based Fingerprinting

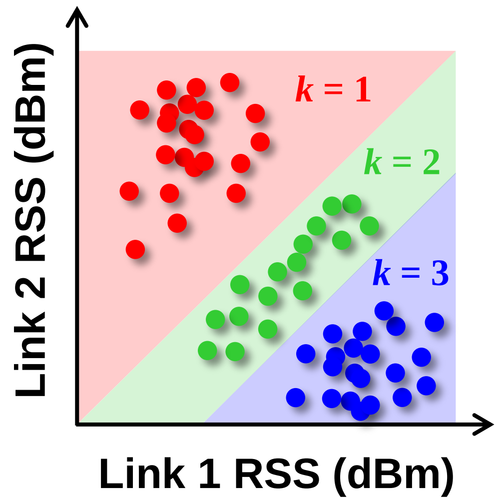


Linear Discriminant Analysis

- RSS measurements with person's presence in each cell is treated as a class/state ***k***
- Each class ***k*** is Multivariate Gaussian with common covariance
- Linear discriminant function:

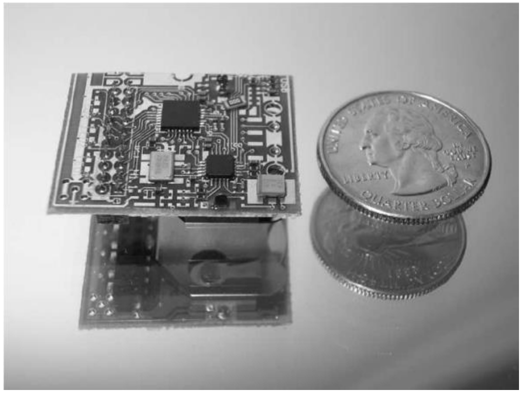
$$\delta_k(x) = x^T \Sigma^{-1} \mu_k - \frac{1}{2} {\mu_k}^T \Sigma^{-1} \mu_k + \log \pi_k$$

$$\hat{y} = argmax_k \delta_k(x)$$



Evaluation Platform

- Hardware: PIP tag
  - Microprocessor: C8051F321
  - Radio chip: CC1100
  - Power: Lithium coin cell battery



- Protocol: Unidirectional heartbeat (Uni-HB)
  - Packet size: 10 bytes
  - Beacon interval: 100 msec

Localization in a cluttered room

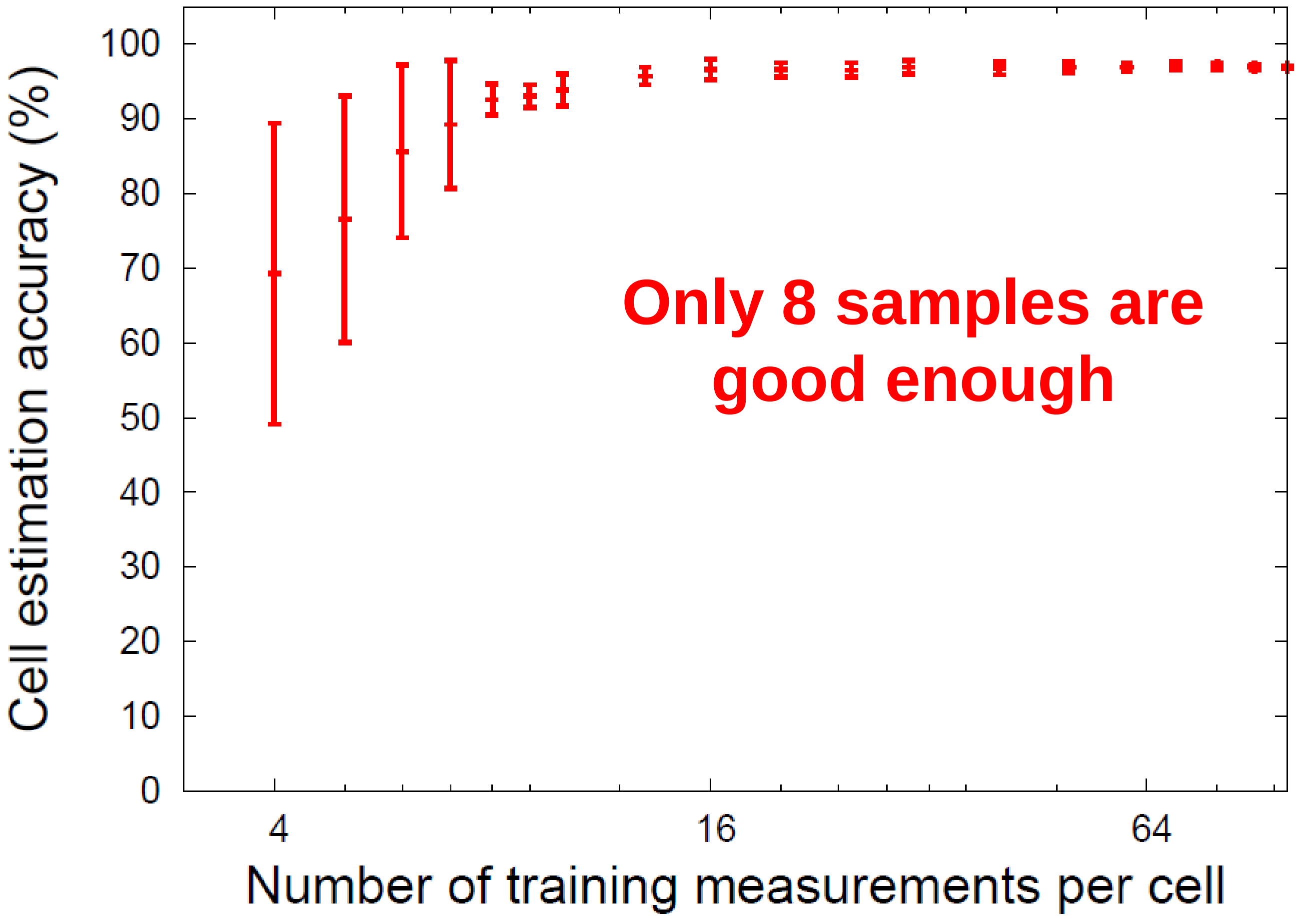
Size: 5 × 8 m

Cell Number: 32

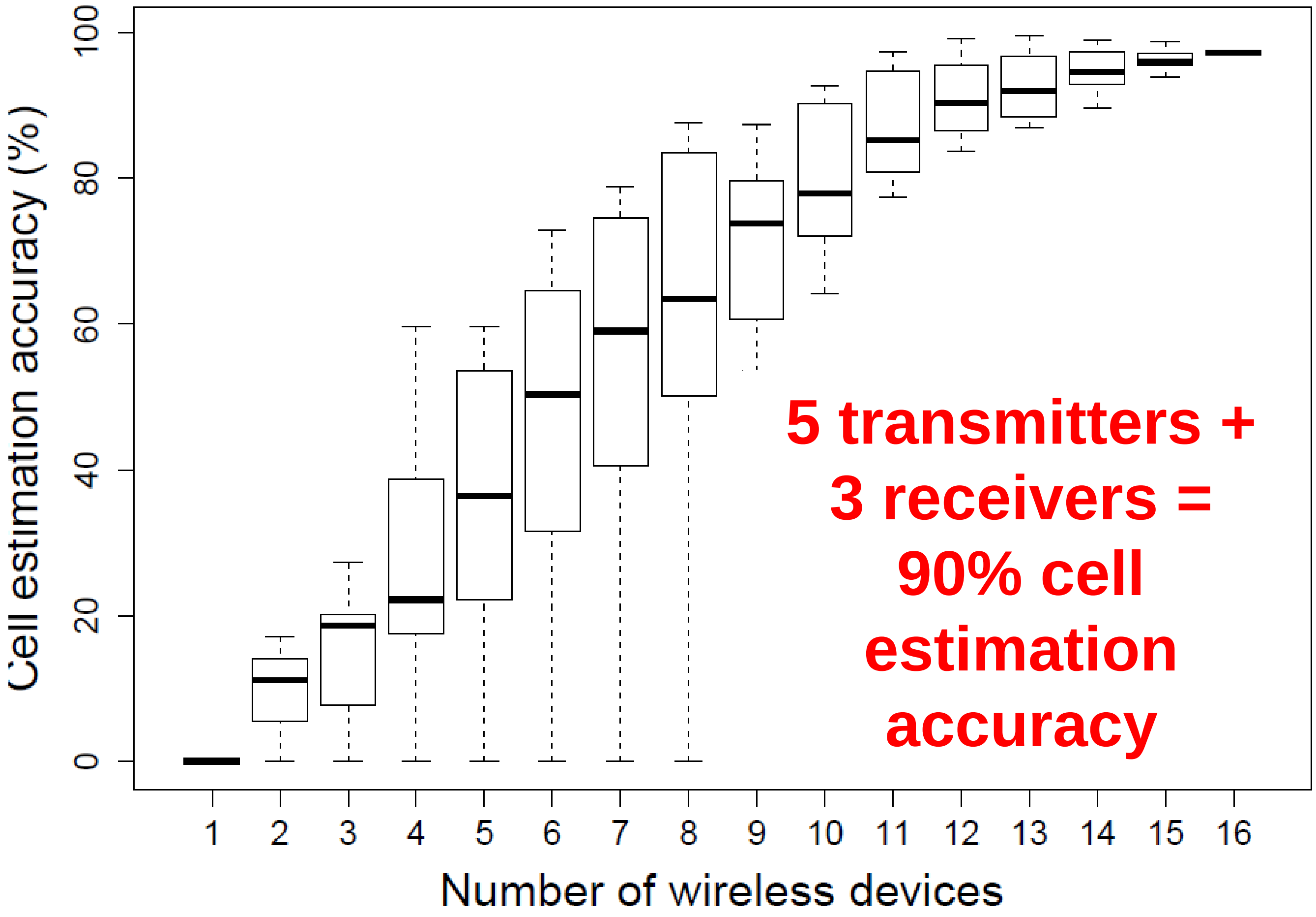
97% cell estimation accuracy (16 devices)

90% Cell estimation accuracy (8 devices)

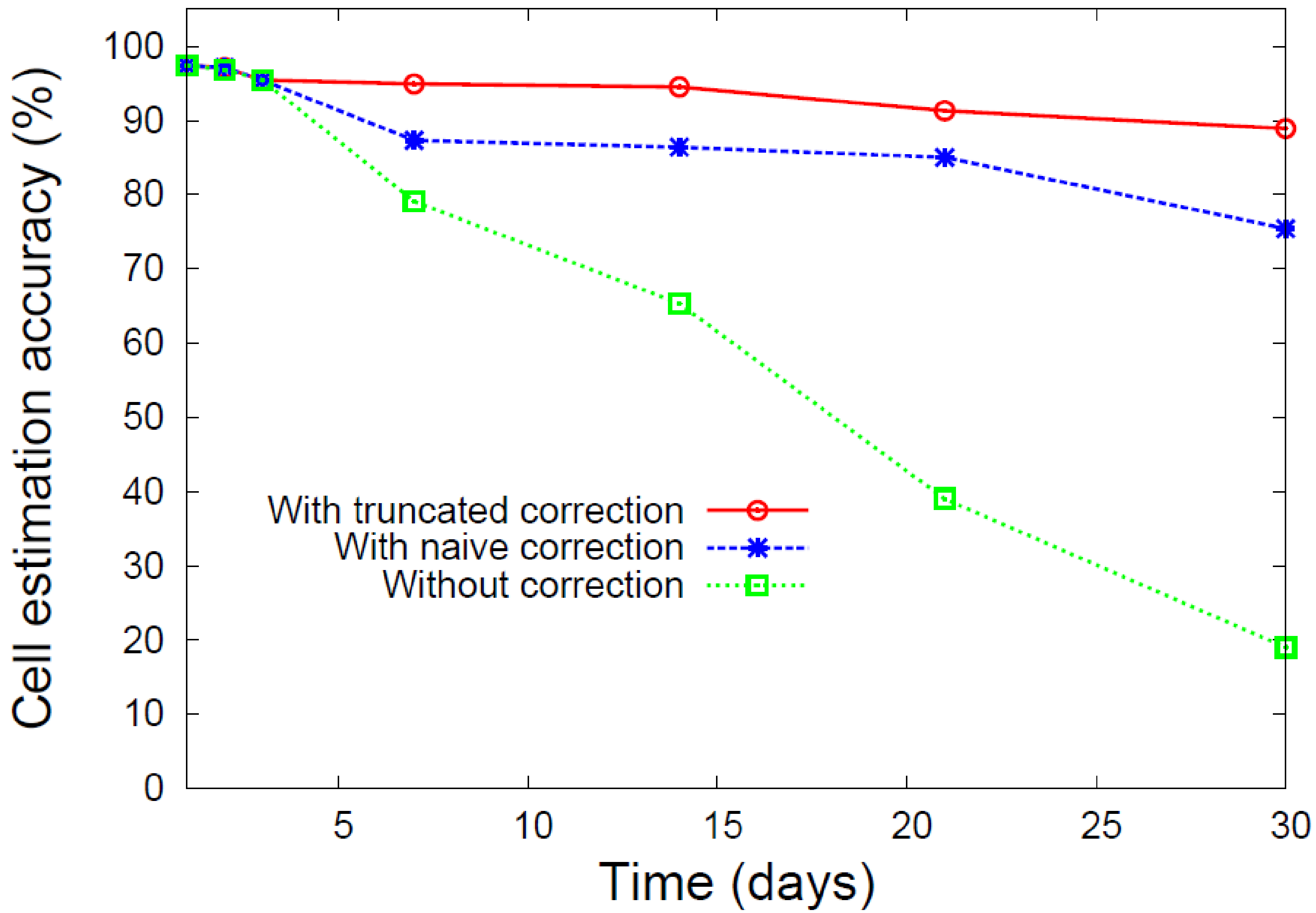
Less training is OK



Having fewer devices is OK



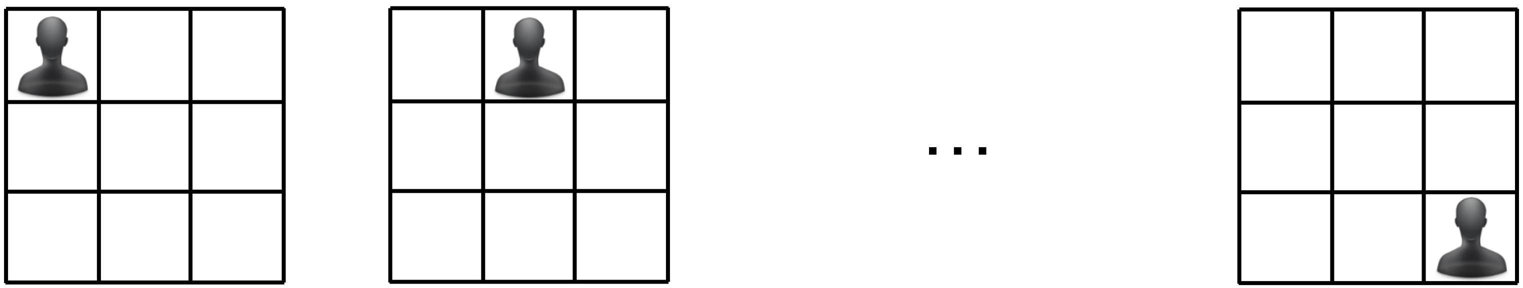
Can we use the same training after 3 months?



Next, let us localize multiple people

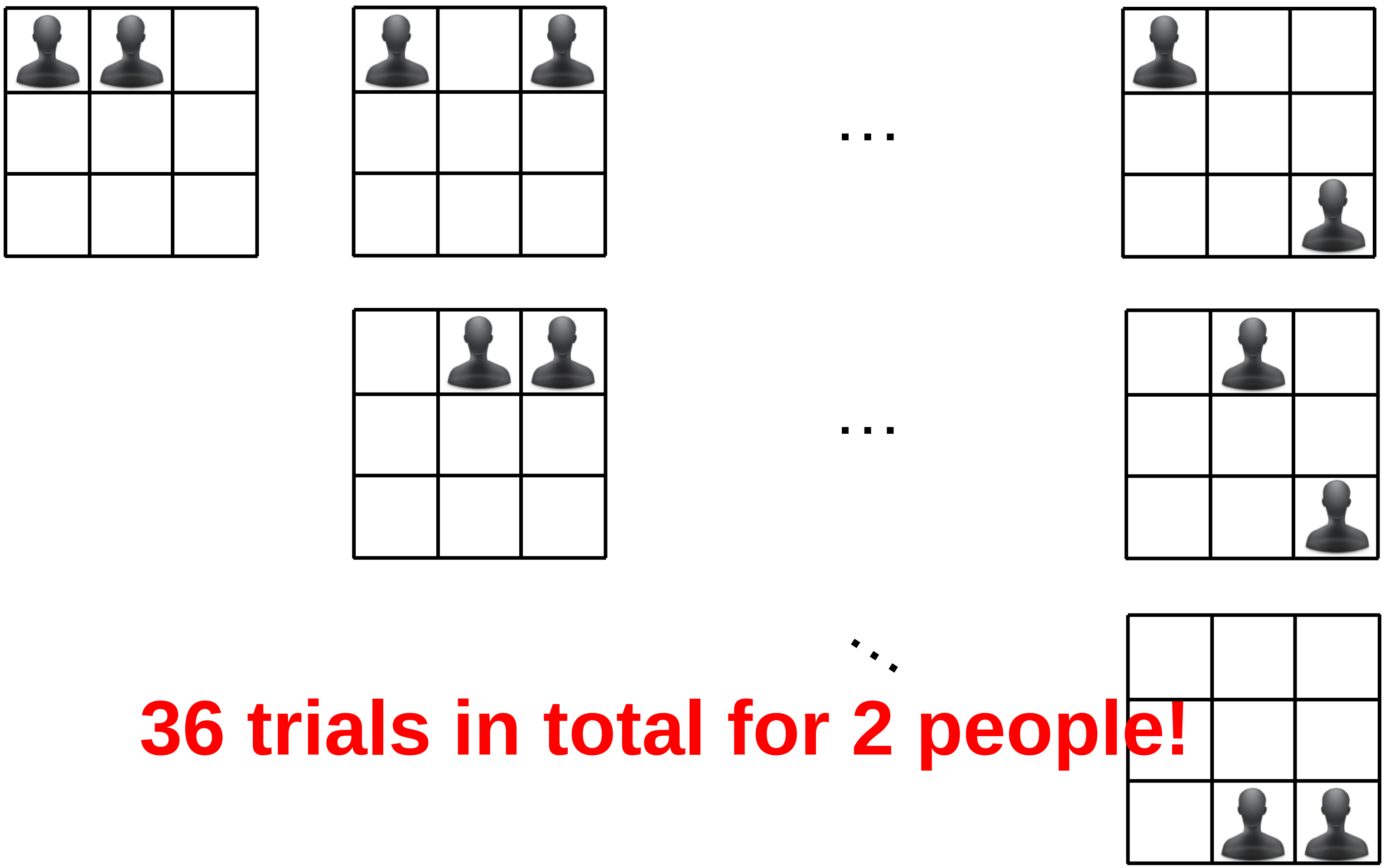
- Challenge: we do NOT want to train all N people with all the combinations at different cells

Fingerprinting 1 person



9 trials in total for 1 person

Fingerprinting 2 people



36 trials in total for 2 people!

Fingerprinting N people

	1 person	2 people	3 people
9 cells	9	36	84
36 cells	36	630	7140
100 cells	100	4950	161700

161700 × 1 min = 112 days

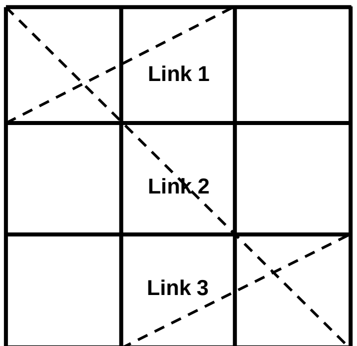
The calibration effort is prohibitive!

Instead,

- Can we use 1 person's training data to localize N people?

- Yes. SCPL has two phases: (i) counting and (2) tracking

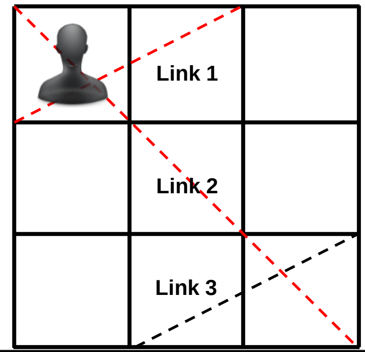
RSS change with people



Link 1 change: 0 dB

Link 2 change: 0 dB

Link 3 change: 0 dB

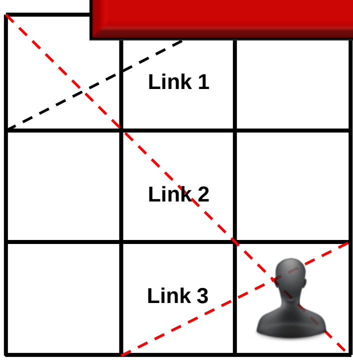


Link 1 change: 4 dB

Link 2 change: 5 dB

Link 3 change: 0 dB

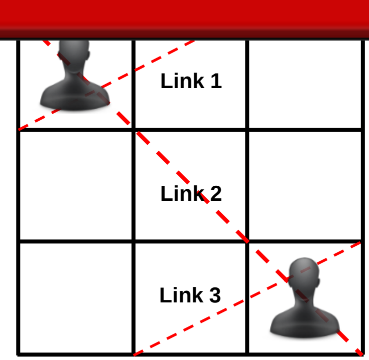
Additive effect on the radio links!



Link 1 change: 0 dB

Link 2 change: 6 dB

Link 3 change: 5 dB



Link 1 change: 4 dB

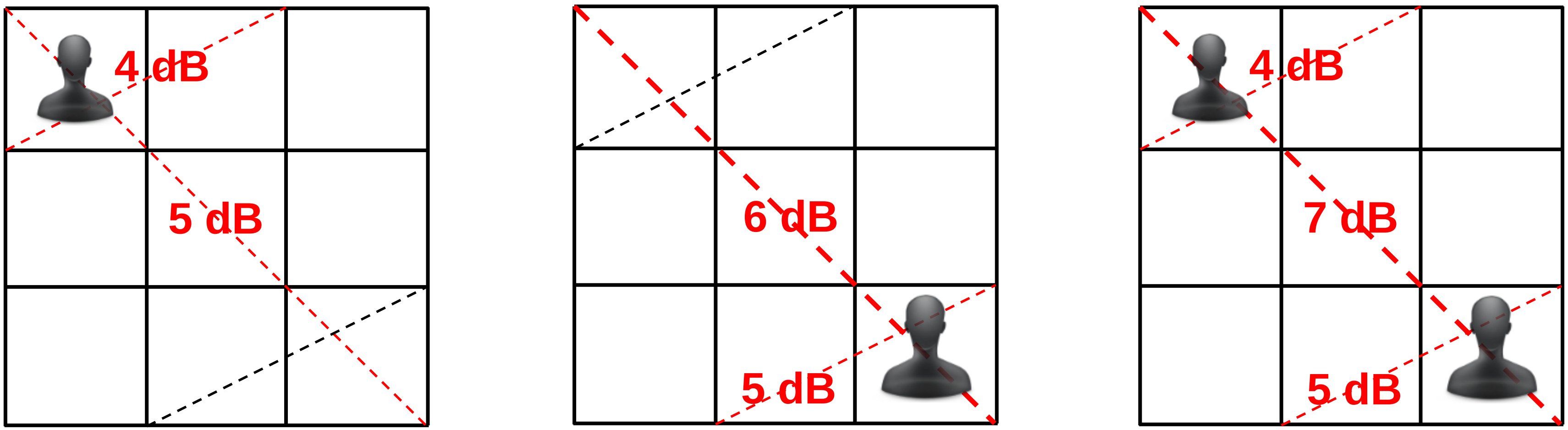
Link 2 change: 7 dB

Link 3 change: 5 dB

So,

- Can we directly infer n from the observed total RSSI change?
- Is it linear?

Nonlinear fading effect!



Shared links observe nonlinear fading effect from multiple people.

4 dB + 0 dB = 4 dB ✓

5 dB + 6 dB = 11 dB ≠ 7 dB X

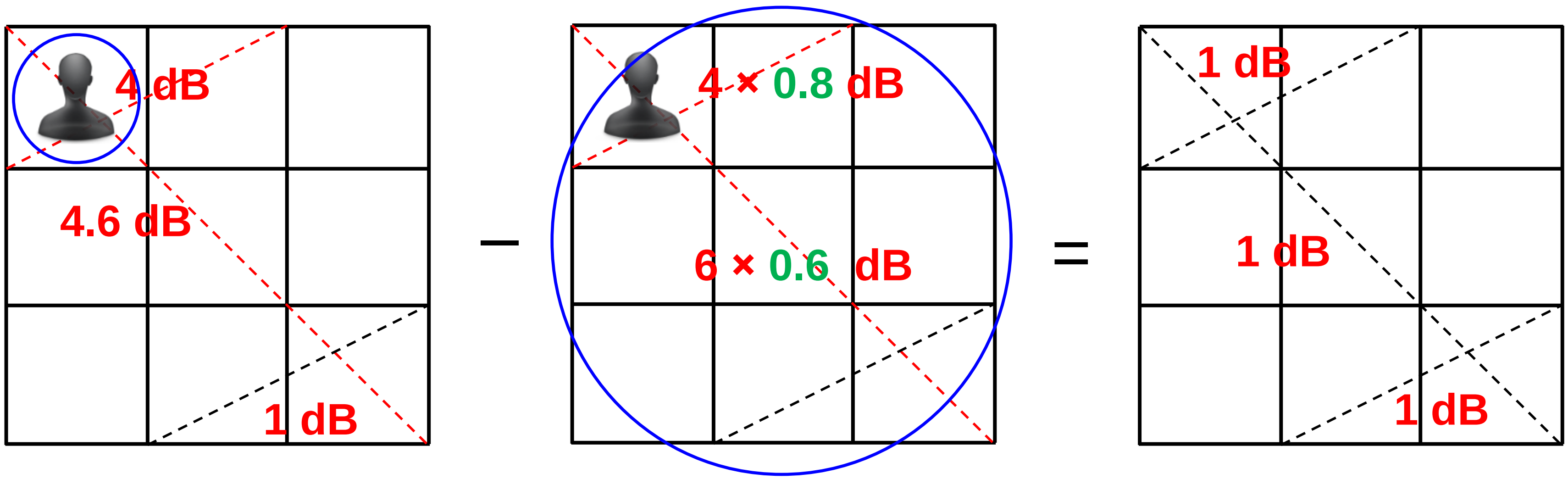
0 dB + 5 dB = 5 dB ✓

Location-Link Correlation

- To mitigate the error caused by this over-subtraction problem, we propose to multiply a location-link correlation coefficient before successive subtracting:

$$\beta_{il} = \frac{h_{ii}^{l}}{\sqrt{\sum_{j=1}^{K} {h_{ij}^{l}}^{2}}} \qquad h_{ij} \leftarrow E\left[\mathcal{D}_{Il}\mathcal{D}_{Jl}\right]$$

Counting Algorithm

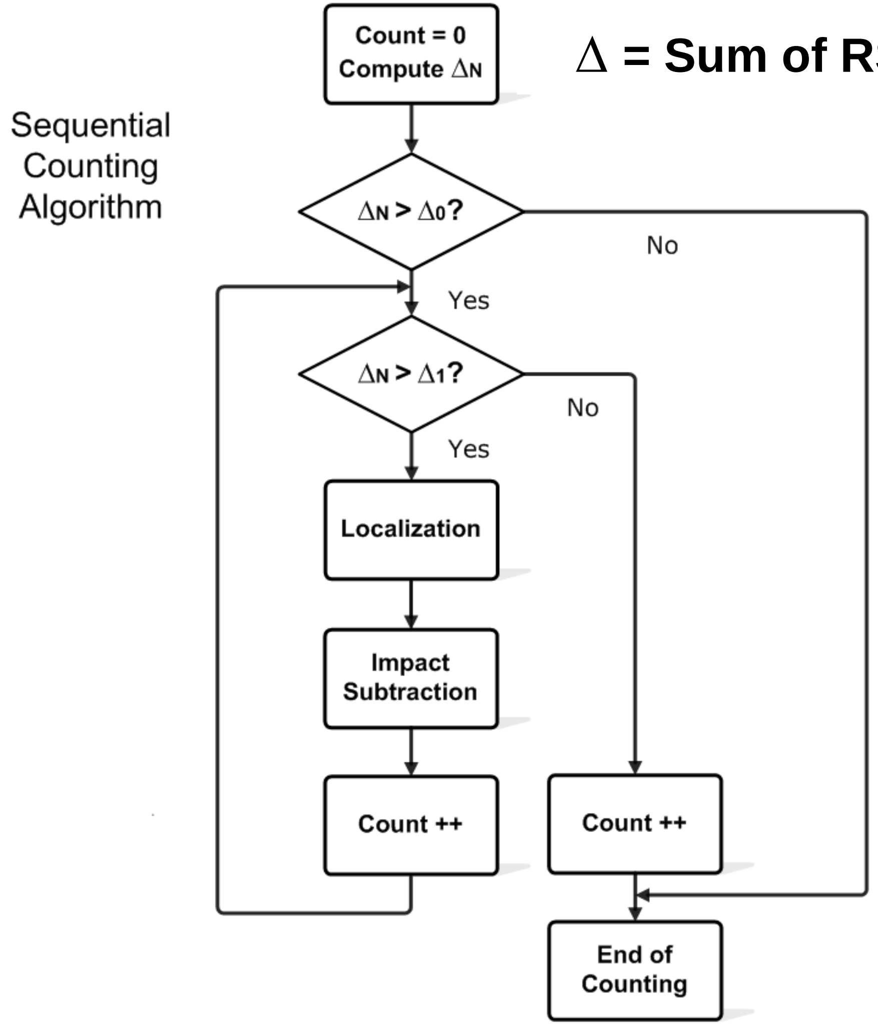


Measurement in 2st round — **Calibration data** — **Measurement In 3rd round**

There are two people in this room.

Sequential Counting (SC) Algorithm

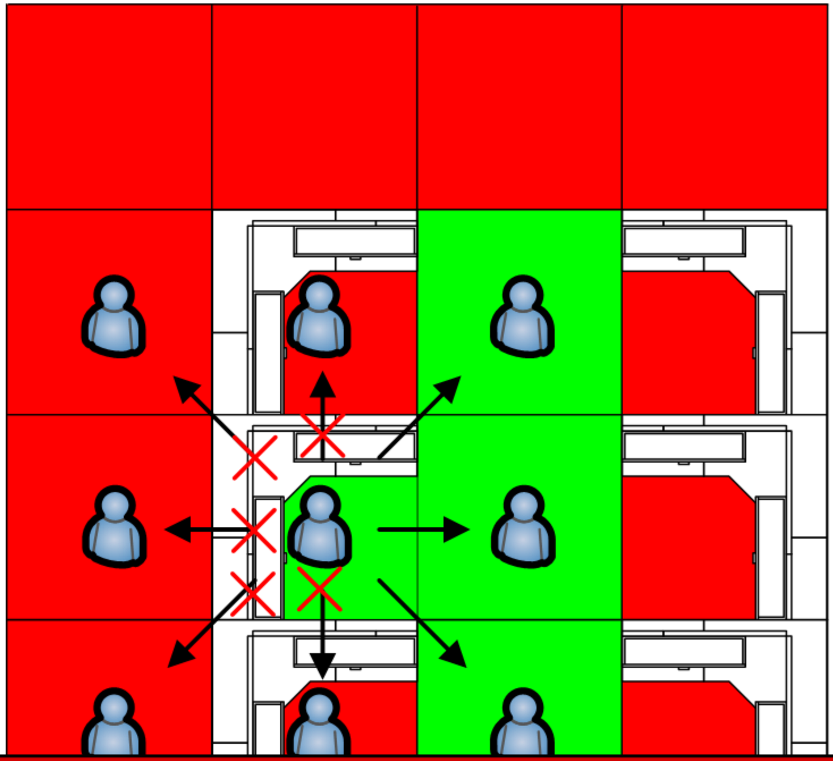
Δ = Sum of RSS change of links



Parallel Localization (PL)

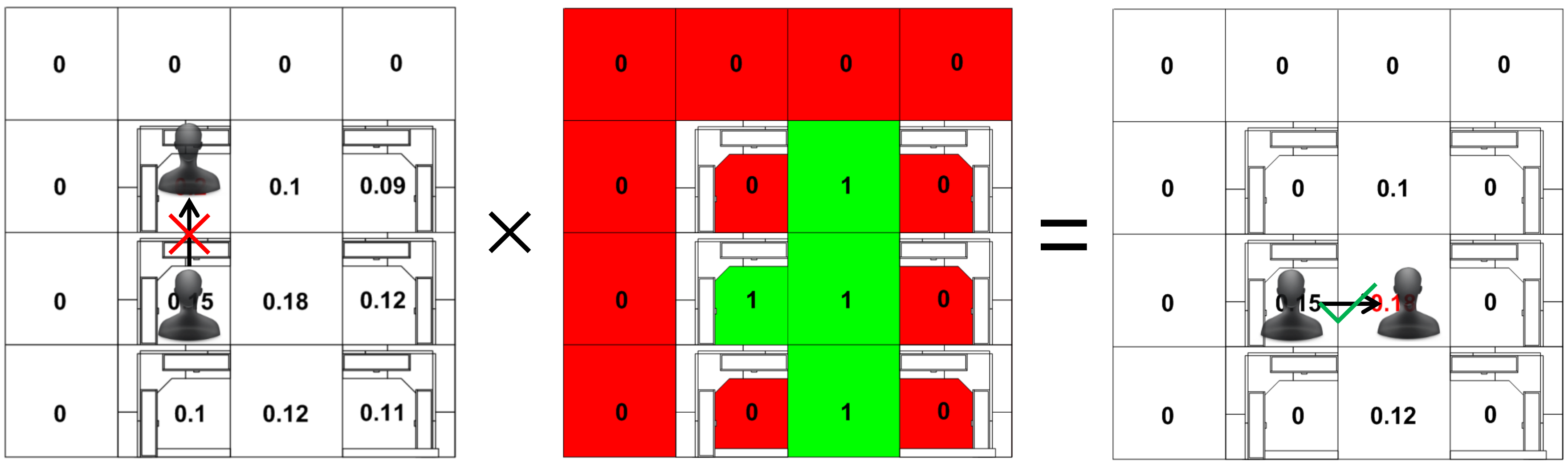
- Cell-based localization
- Trajectory-assisted localization
  - Improve accuracy by using human mobility constraints

Mobility makes localization easier



In a building, your next step is constrained
by walking speed, cubicles, walls, etc.

Trajectory-based Localization



Data likelihood map Trajectory ring filter Refined likelihood map

Indoor mobility constraints can help improve localization accuracy.

Parallel Localization (SL) Algorithm

- **Single subject localization**

$$V_j(t) = \underset{q_1, q_2, \ldots, q_{t-1}}{\operatorname{argmax}} P(q_1 q_2 \ldots q_t = j, O_1 O_2 \ldots O_t | T, \delta)$$

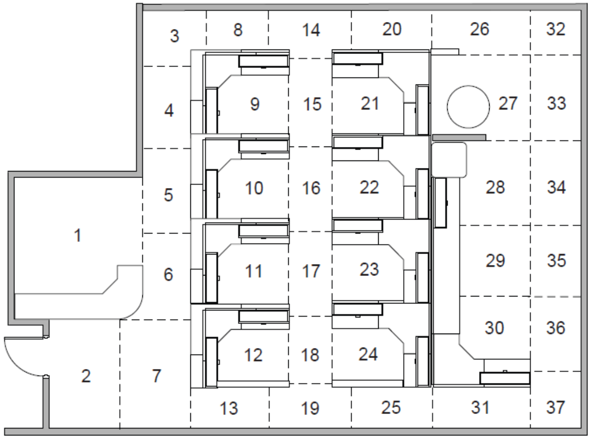
- **Multiple subjects localization**

$$\texttt{ViterbiScore} = F_j = \sum_{i=1}^{C} \delta_{q_t^i}(O_t) T_{q_{t-1}^i q_t^i}$$

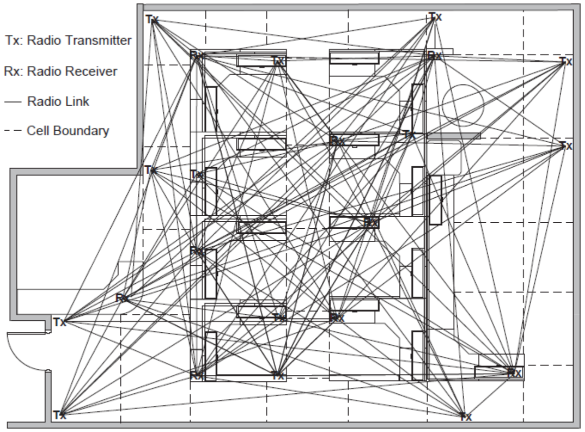
$$\Pi \leftarrow \text{is the set of all the possible permutations of } \binom{K}{C}$$
$$Q_i \leftarrow \operatorname{argmax}_{j \in \Pi} \texttt{ViterbiScore}(Q_{i-1}, Q_j, \delta_{1:K}(O_i), T)$$

Testing Environment

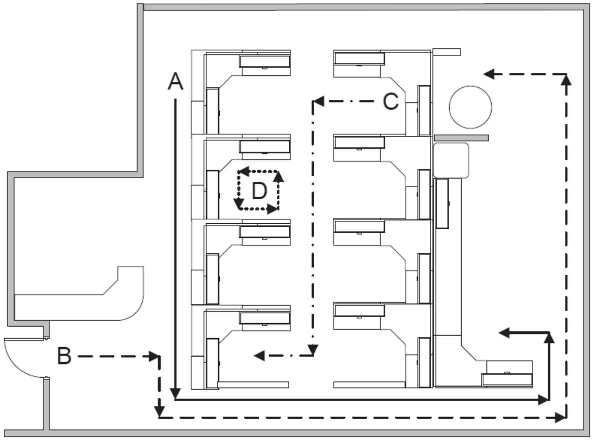
Total size: 10 × 15 m



37 cells of cubicles and aisle segments

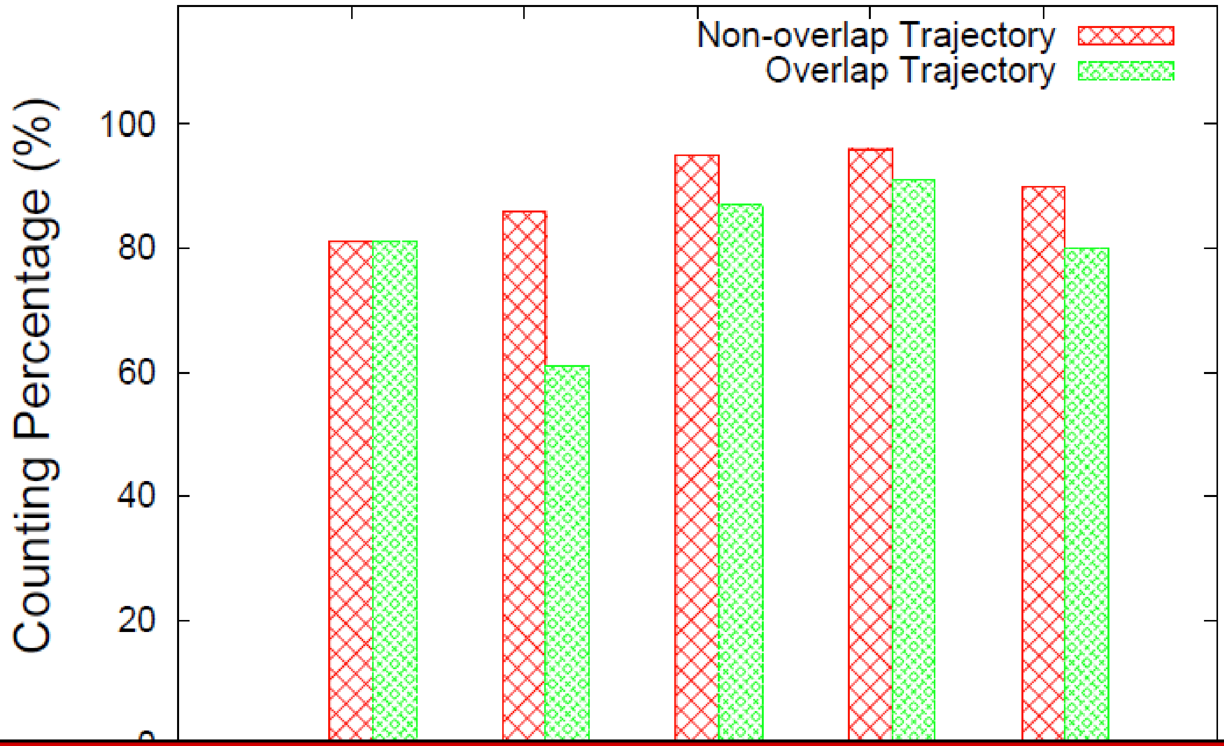


13 transmitters and 9 receivers



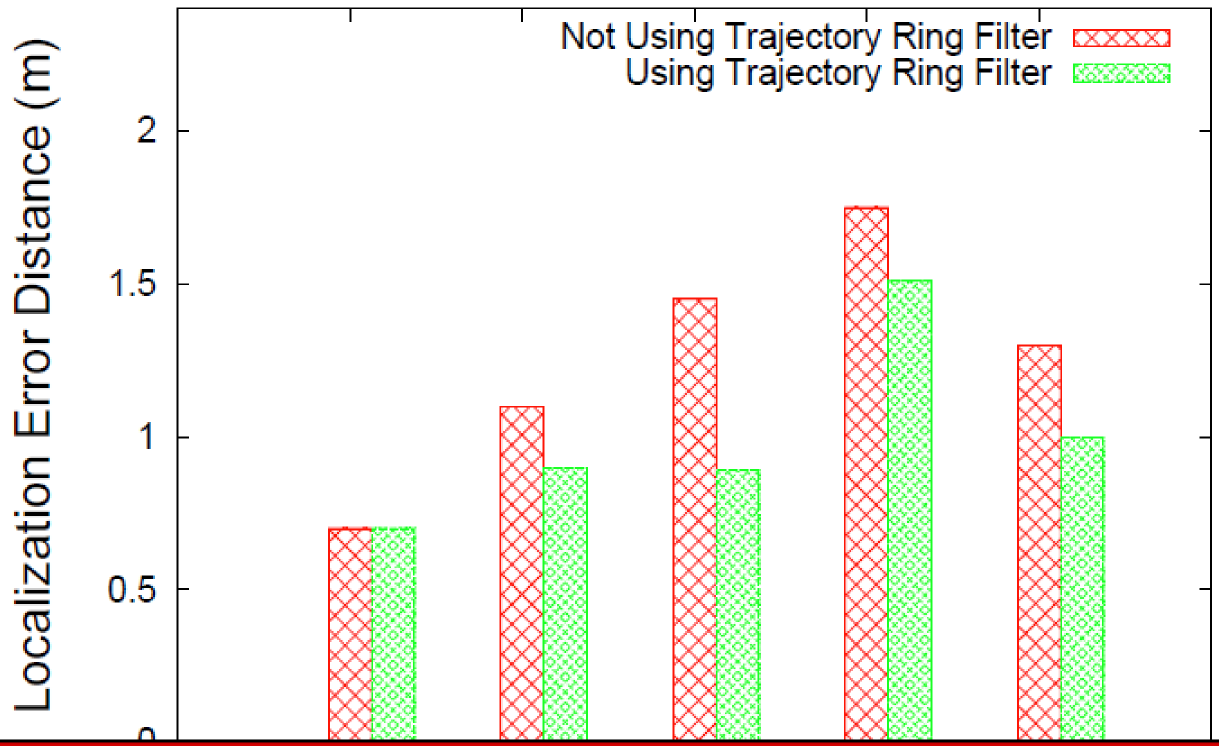
Test paths with partial overlap

Counting Results



We achieve above 85% counting accuracy when no trajectories are overlapped.

Localization Results



Trajectory ring filter achieve 1-meter localization accuracy and improve 30% from the baseline.

Lessons learned

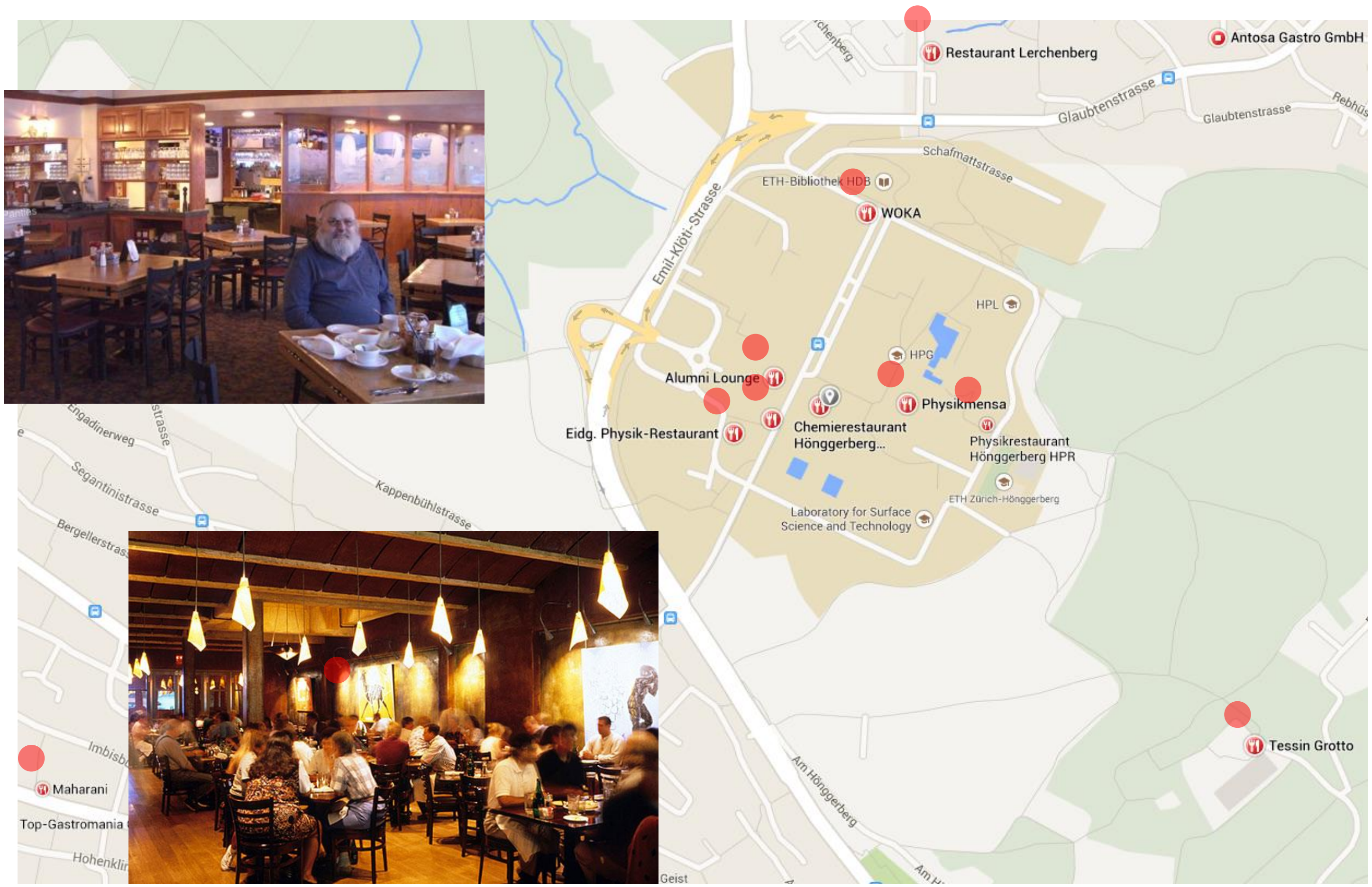
- Calibration data collected from one subject can be used to count and localize multiple subjects.
- Though indoor spaces have complex radio propagation characteristics, the increased mobility constraints can be leveraged to improve tracking accuracy.

Crowd++

Unsupervised Speaker Counting on Smartphones: speaker count

C. Xu, S. Li, G. Liu, Y. Zhang, E. Miluzzo, Y. Chen, J. Li, B. Firner. Crowd++: Unsupervised Speaker Count with Smartphones. In ACM UbiComp, 2013

Scene 1: Dinner time, where to go?



Scene 2: Is your kid social?






Scene 3: Which class is engaging?

Speaker count

- Dinner time, where to go?
  - Find the place where has most people talking!
- Is your kid social?
  - Find how many (different) people they talked with!
- Which class is more attractive?
  - Check how many students ask you questions!

Microphone + microcomputer

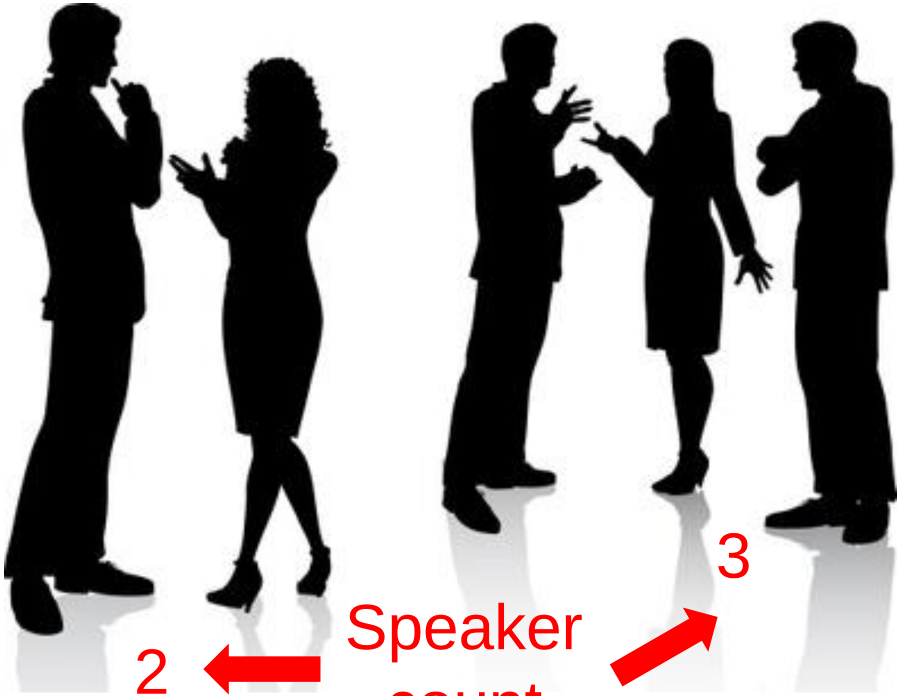
Conversation contexts

Speech recognition

Family life

Bob Alice

Speaker identification



Stressful

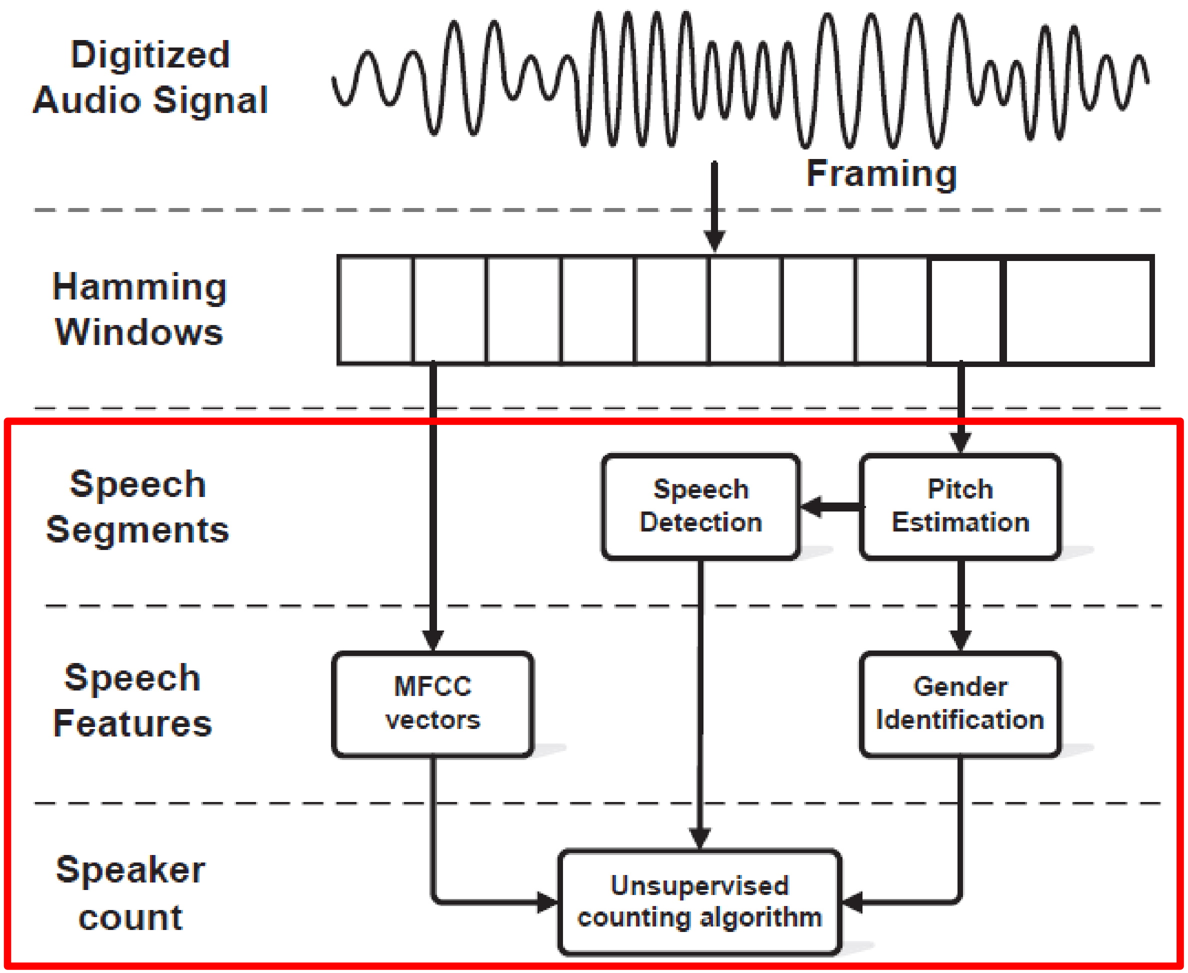
Emotion detection

3

2

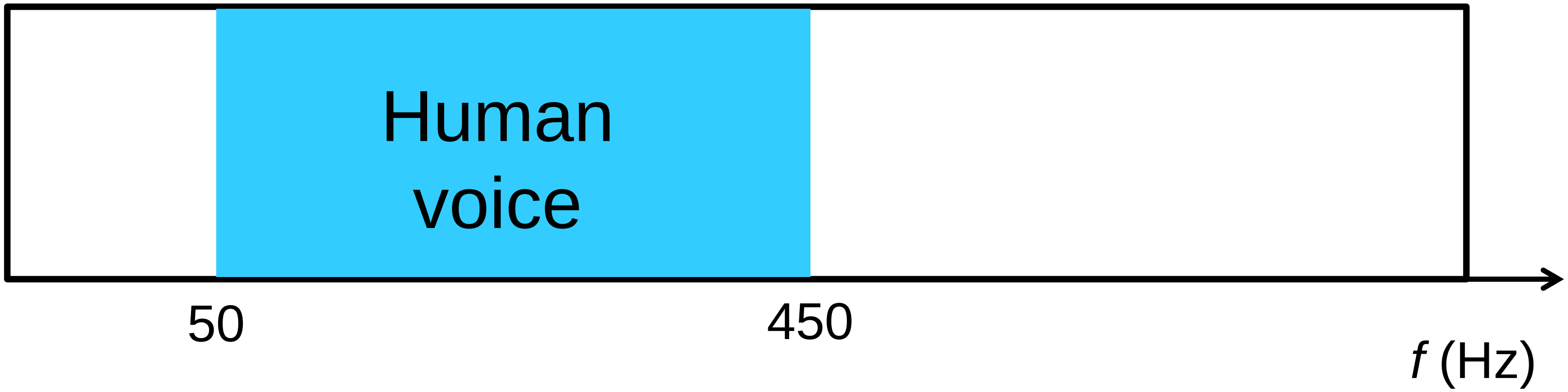
Speaker count

Overview


Digitized Audio Signal
Framing
Hamming Windows
Speech Segments
Speech Detection
Pitch Estimation
Speech Features
MFCC vectors
Gender Identification
Speaker count
Unsupervised counting algorithm

Speech detection

- Pitch-based filter
  - Determined by the vibratory frequency of the vocal folds
  - Human voice statistics: spans from 50 Hz to 450 Hz



Speaker features

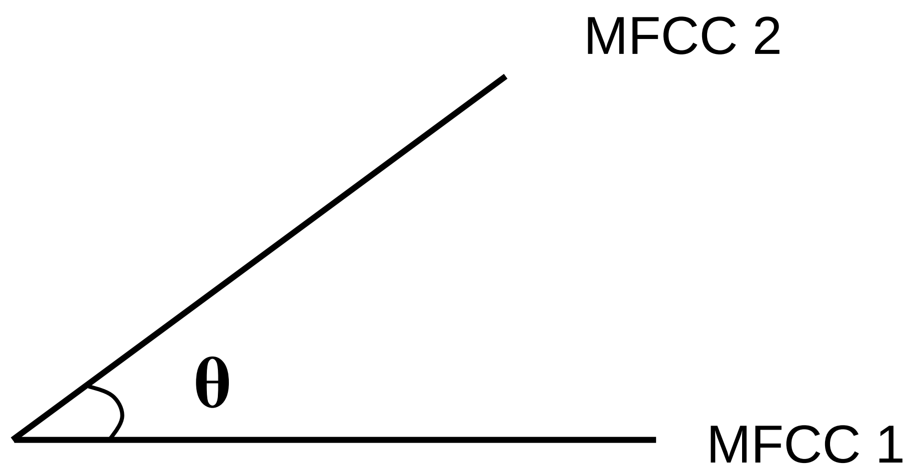
- MFCC
  - Speaker identification/verification
    - Alice or Bob, or else?
  - Emotion/stress sensing
    - Happy, or sad, stressful, or fear, or anger?
  - Speaker counting
    - No prior information needed

Supervised

Unsupervised

Speaker features

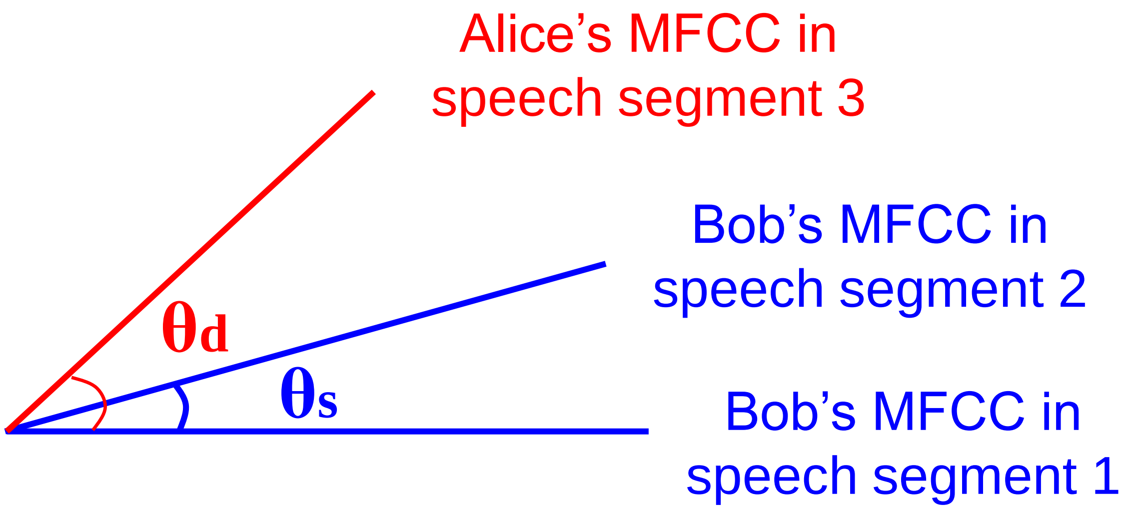
- MFCC + cosine similarity distance metric



We use the angle θ to capture the distance between speech segments.

Speaker features

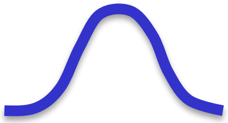
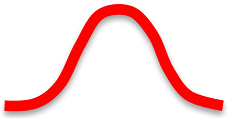
❑ MFCC + cosine similarity distance metric



$\theta_d > \theta_s$

Speaker features

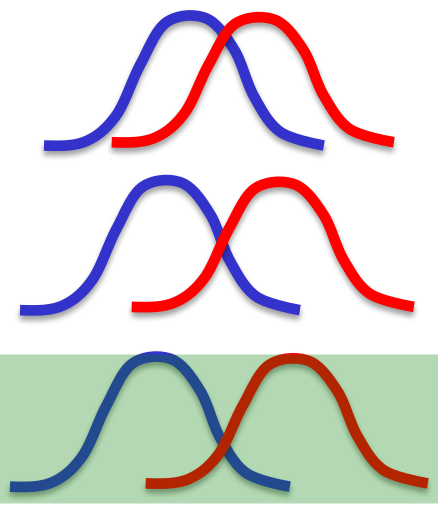
- MFCC + cosine similarity distance metric

 **histogram of θ_s**  **histogram of θ_d**

1 second
speech segment

2-second
speech segment

3-second
speech segment

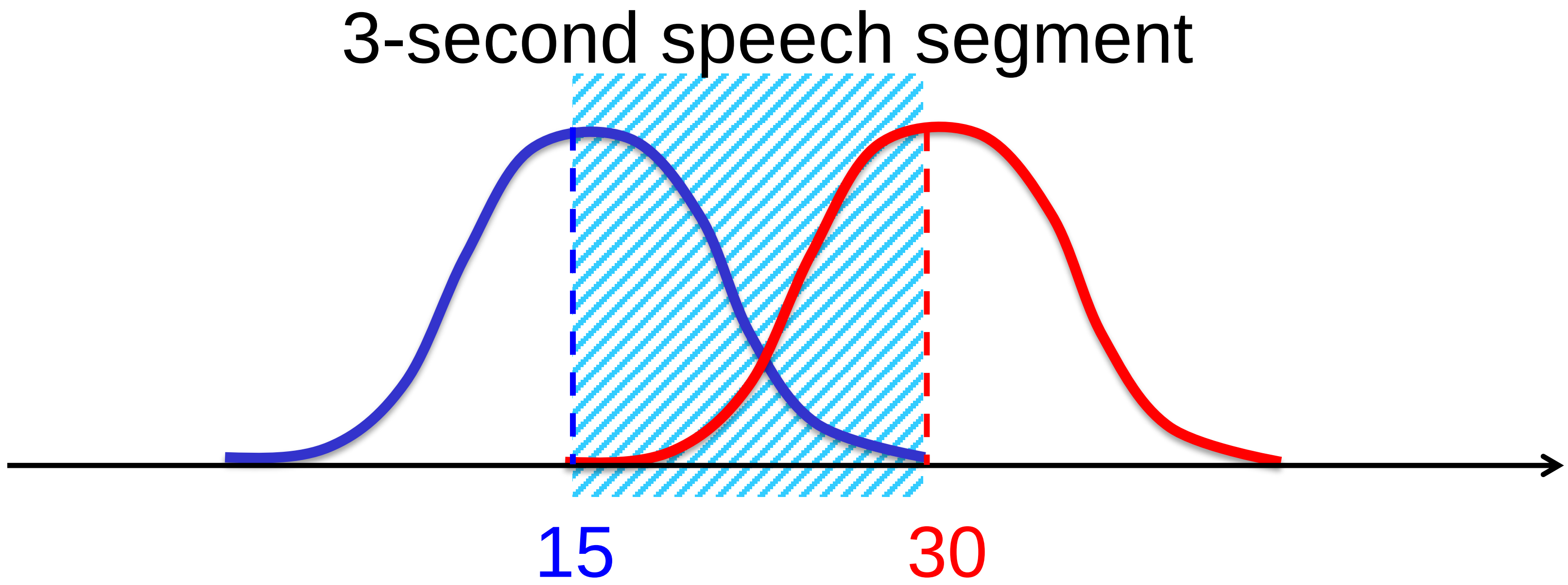


.
.

10-second utterance is not common in conversation!

Speaker features

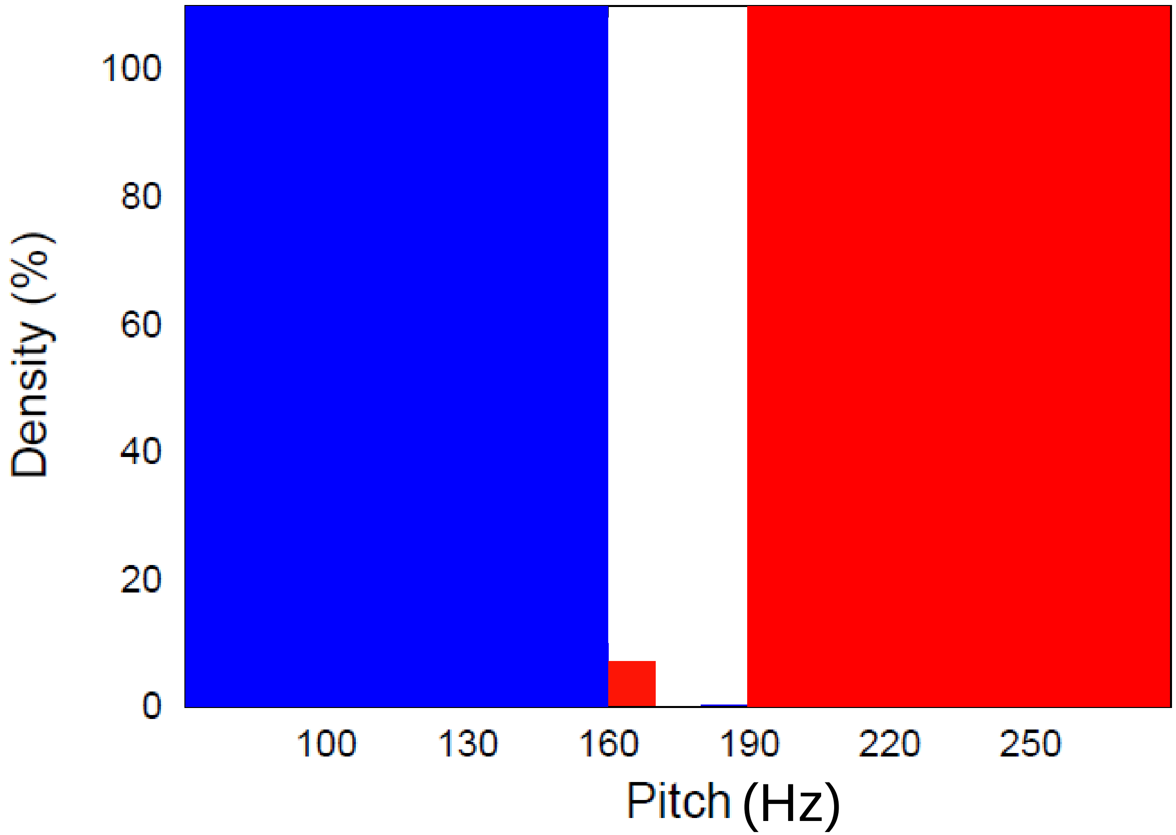
- MFCC + cosine similarity distance metric



Thresholds trade-off the sensitivity to admitting new speaker, as well as filtering overlap/silence.

Speaker features

- Pitch + gender statistics



Same speaker or not?

IF MFCC cosine similarity score < 15

AND

Pitch indicates they are same gender

Same speaker

ELSEIF MFCC cosine similarity score > 30

OR

Pitch indicates they are different genders

Different speakers

ELSE

Not sure

Evaluation through crowdsourcing

- 120 users from university and industry contribute 109 audio clips of 1034 minutes in total.

Private indoor | Public indoor | Outdoor

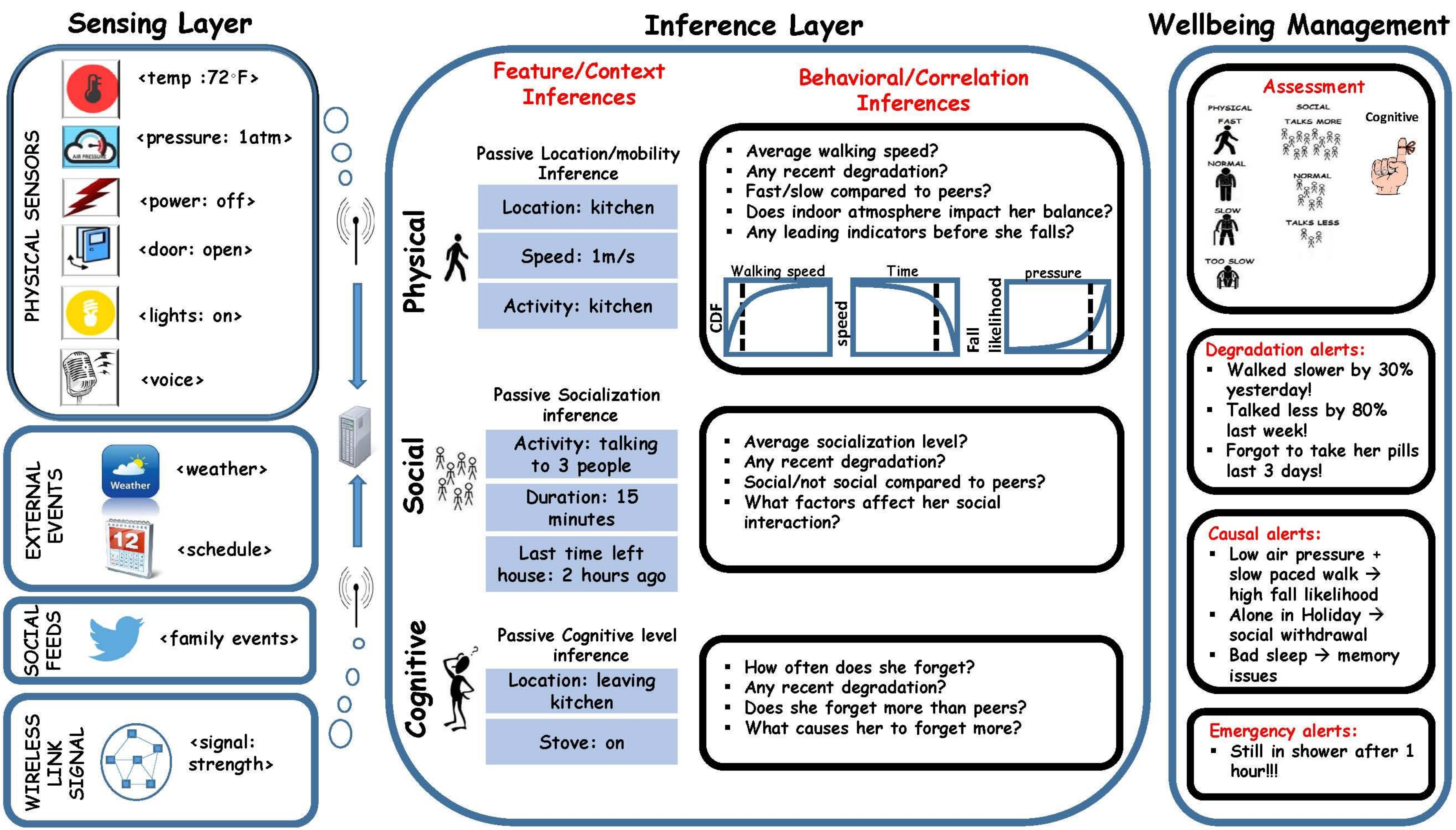
Crowdsourcing results

	Sample number	Error count distance
Private indoor	**40**	**1.07**
Public indoor	**44**	**1.35**
Outdoor	**25**	**1.83**

Lessons learned

- Accuracies: private indoor > public indoor > outdoor
- We need low-cost noise cancellation technique to improve the accuracy

Ongoing work – Elder care with SCPL + Crowd++ + many more



Questions & Answers