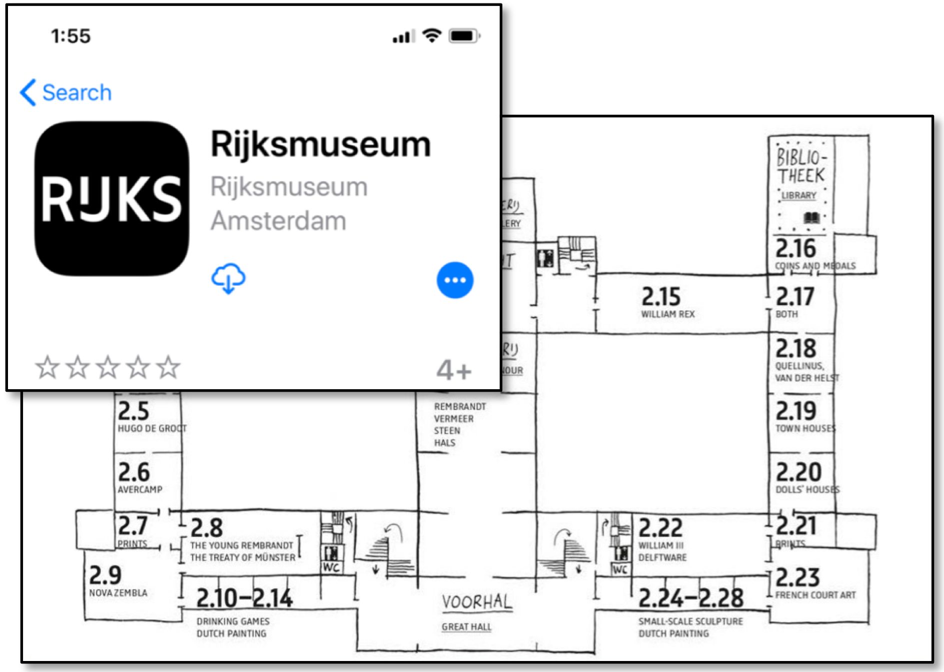
# Deep Room Recognition Using Inaudible Echos

Qun Song [1,2]  Chaojie Gu[2]  Rui Tan[2]

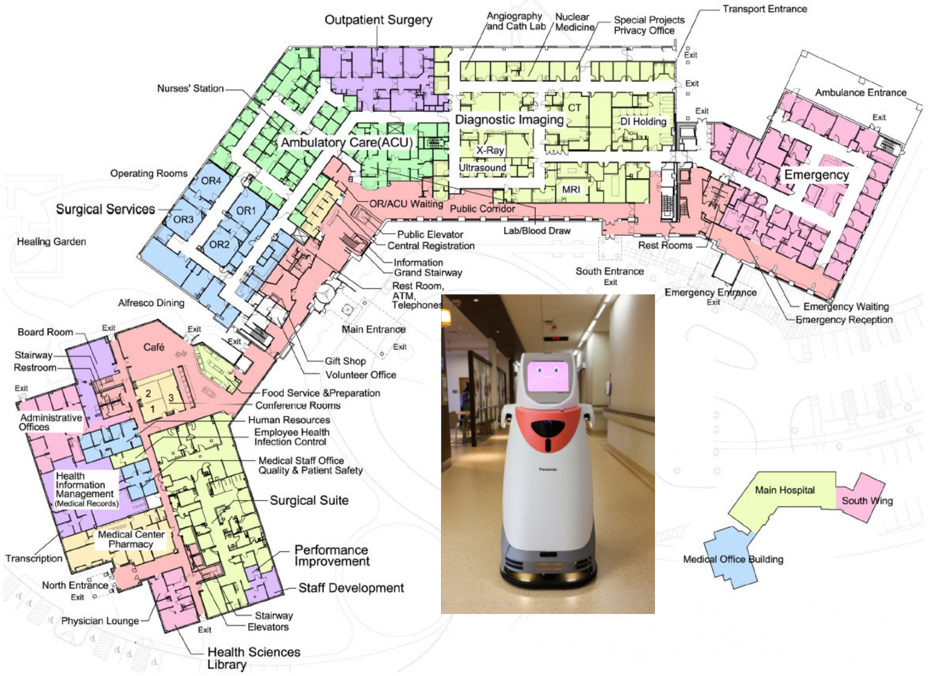
[1]Energy Research Institute, Interdisciplinary Graduate School
[2]School of Computer Science and Engineering

Nanyang Technological University

# Room-level Localization




Museum with many exhibition chambers




Autonomous delivery robot in
Changi General Hospital, Singapore

- Useful in a range of applications
  - Automated multimedia guide in a museum
  - Robot localization & patient/newborn tracking in a hospital

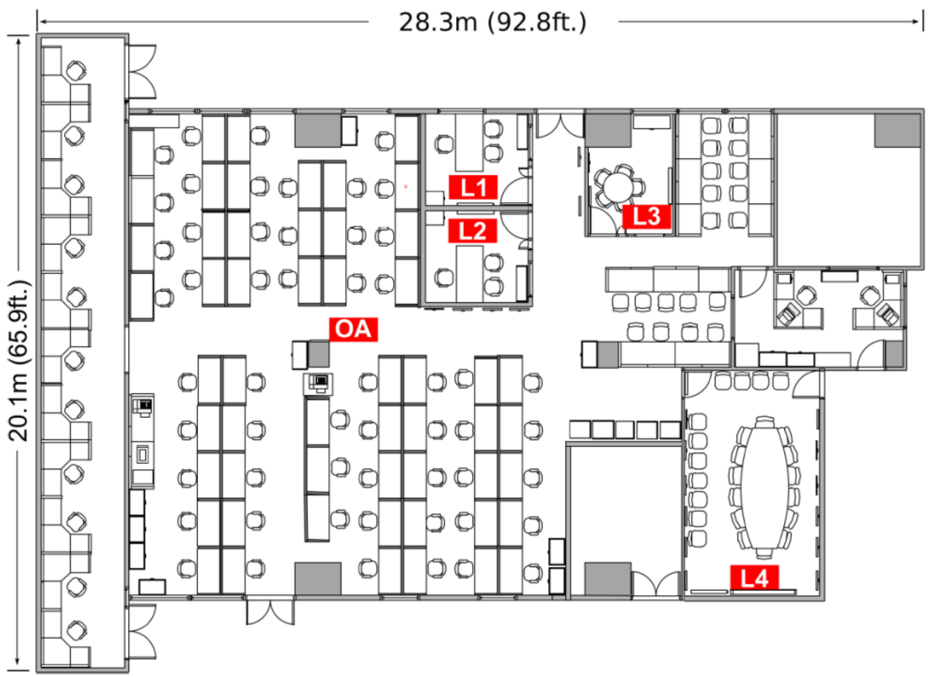
# Objective

- **Reliable** room-level localization using phone/wearable built-in audio system only
  - Infrastructure-free
  - No add-on hardware
- **Practical**
  - Designer: effortless training data collection
  - End users: download and use
- **Privacy-preserving**
  - Very short audio recording

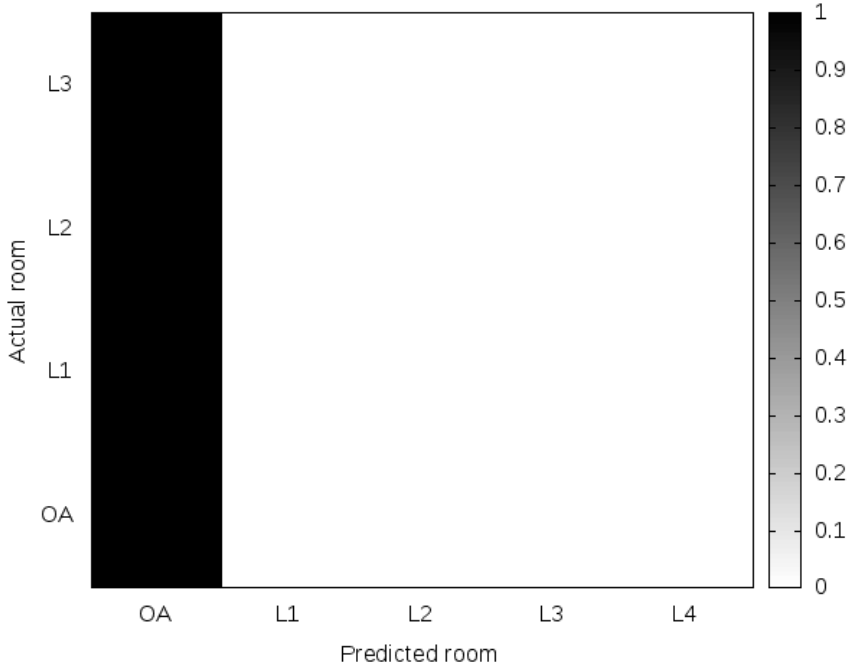
# Related Work

- Passive audio sensing
  - **SurroundSense** [MobiCom'09], **Batphone** [MobiSys'11]
    *Susceptible to interference, privacy breaching (10s recording)*
- Active audio sensing
  - **RoomSense** [AH'13]
    *Uses full-spectrum audio, susceptible to foreground sound*
- Semantic localization
  - Backpack, drawer, restroom, elevator, etc
    *Recognize context, rather than location*

# Susceptibility of Passive Sensing




Laboratory floor plan




Confusion matrix

- Batphone [MobiSys’11]
  - Install on an iPhone 6s from Apple’s App Store
  - Quiet environment: down to **40%** accuracy
  - Ambient music during testing: **0%** accuracy for L1 to L4

# Outline

- Motivation
- **Measurement**
- Approach & Evaluation
- Conclusion

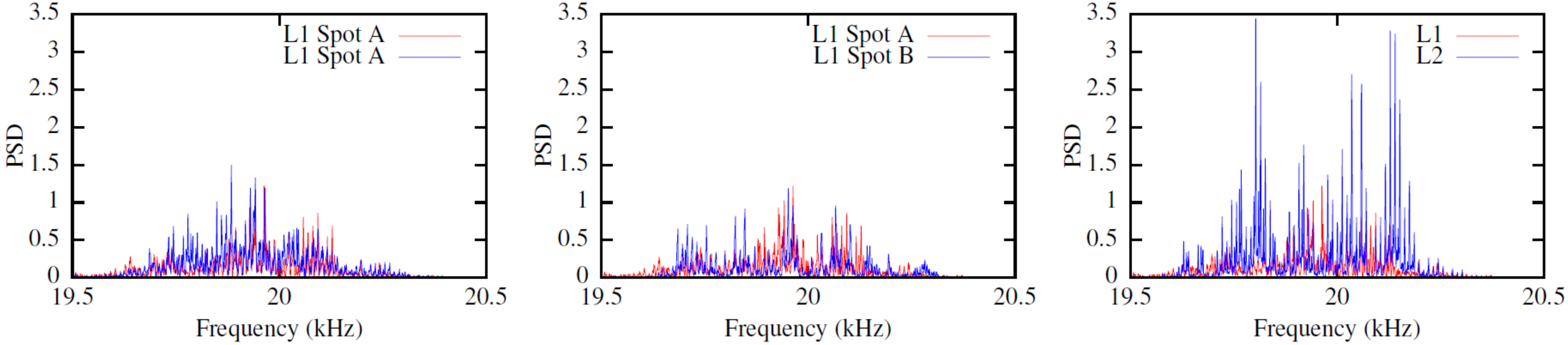
# Probe Signal

- Those used in existing studies
  - Sine sweep, maximum length sequence (MLS), multi-frequency chirp
    *Audible (annoying), wide-band (susceptible to foreground sounds)*

- Short-time single-frequency chirp
  - 2ms
    *Echos from objects >34cm away won't mix with chirp*
  - 20kHz
    *Inaudible, different from man-made sounds*
  - Challenge: limited information carried by echos

# Room's Response

# Frequency Analysis



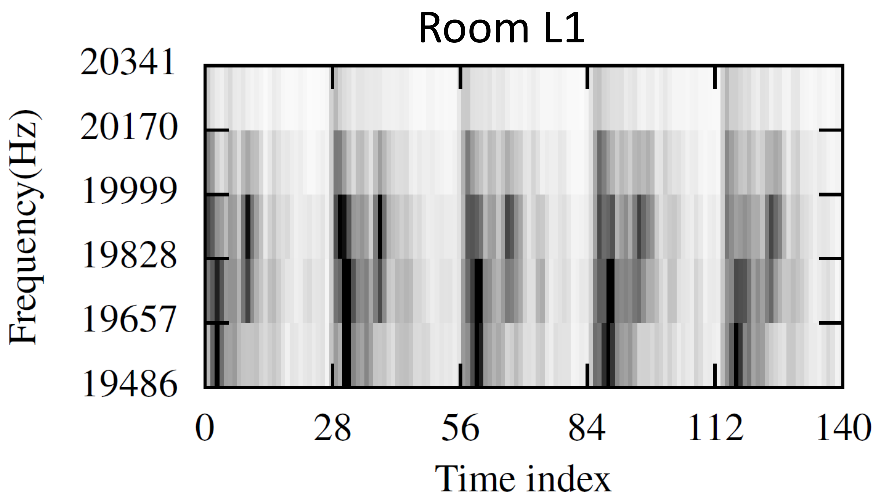

Same room, same spot

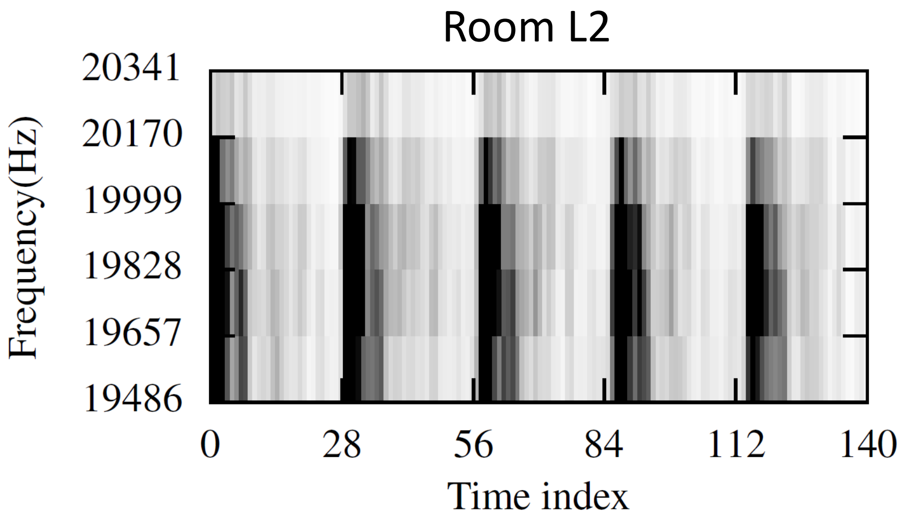
Same room, diff spots

Different rooms

- L1 and L2 have the same size and furniture
  - A room gives stable frequency response
  - Different rooms respond differently

# Time-Frequency Analysis




- Spectrogram
  - Each room has stable spectrogram
  - Perceptible differences for different rooms

# Outline

- Motivation
- Measurement
- **Approach & Evaluation**
- Conclusion

# Candidate Designs

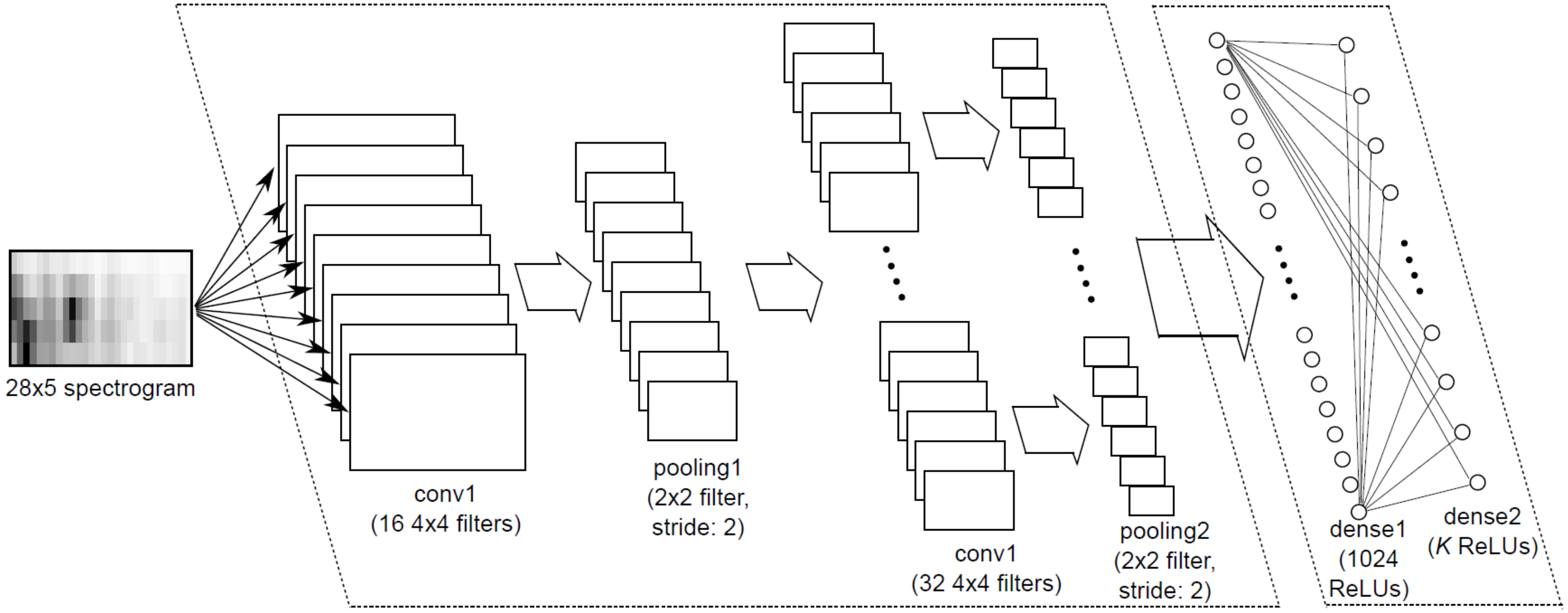
- Existing systems use “shallow” learning (SVM)
  - Manually engineered features
  - Ineffective in addressing subtle differences
- Deep learning
  - Automates feature extraction

Four candidate designs

| | PSD | Spectrogram |
|---|---|---|
| **DNN** | Design 1 | Design 2 |
| **CNN** | Design 3 | Design 4 |

# Data Format and Deep Model

- Google TensorFlow
  - DNN: 2 hidden layers, each with 256 ReLUs
  - CNN:




Test accuracy in classifying 22 rooms

| | PSD | Spectrogram |
|---|---|---|
| DNN | 19% | 80% |
| CNN | 33% | 99% |

# Example Room Types

(a) Bedroom (b) Museum hall (c) Visitor office L1 (d) Lab open area (e) Meeting room L4

Examples of several room types

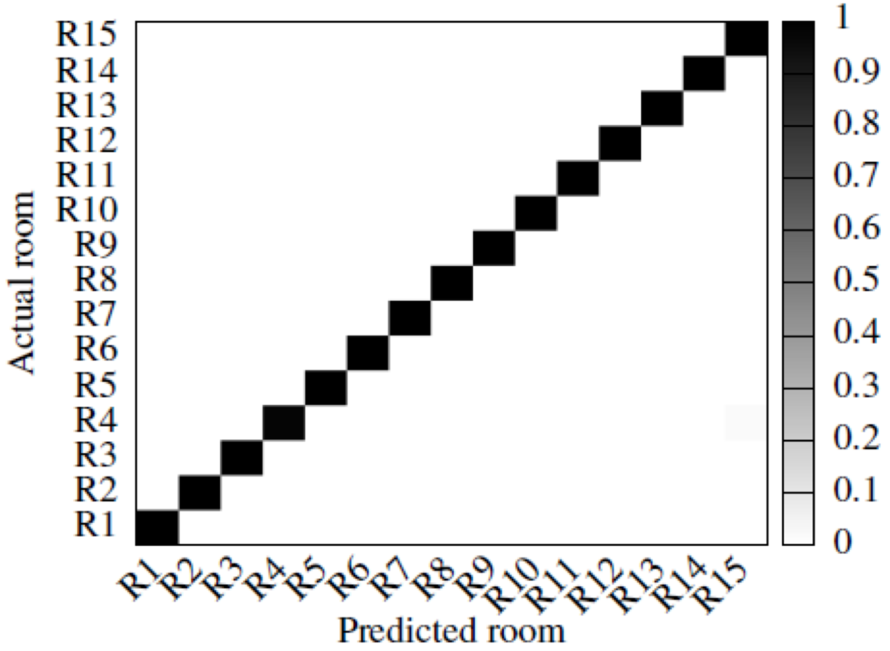
(a) Teaching room 1 (b) Teaching room 2 (c) Teaching room 3 (d) Teaching room 4 (e) Teaching room 5

Examples of similar rooms

# Robustness to Foreground Sound

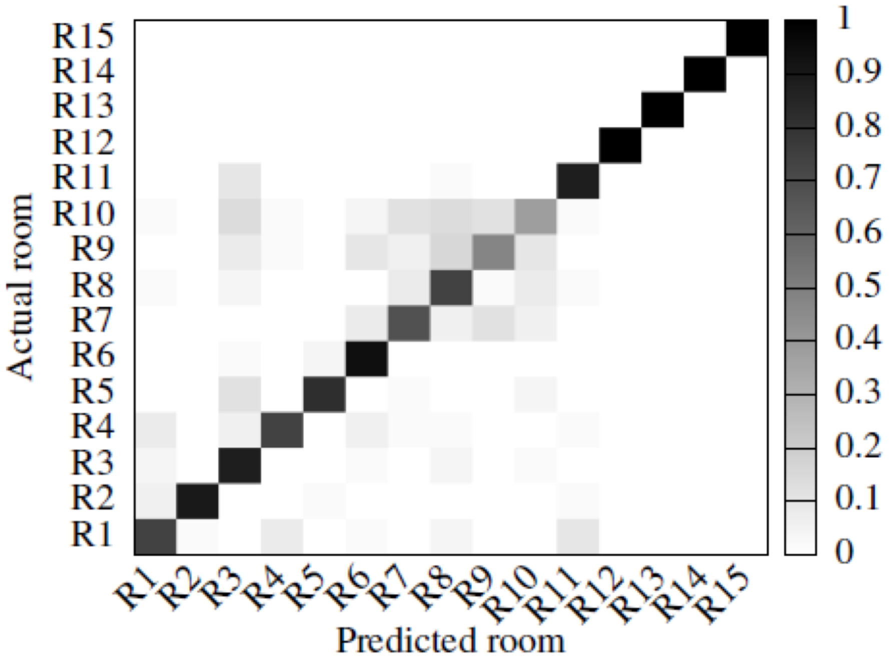
- Test our approach in rooms R1 – R15

Confusion matrix (quiet rooms)




100%

Confusion matrix (play Youtube)




81%

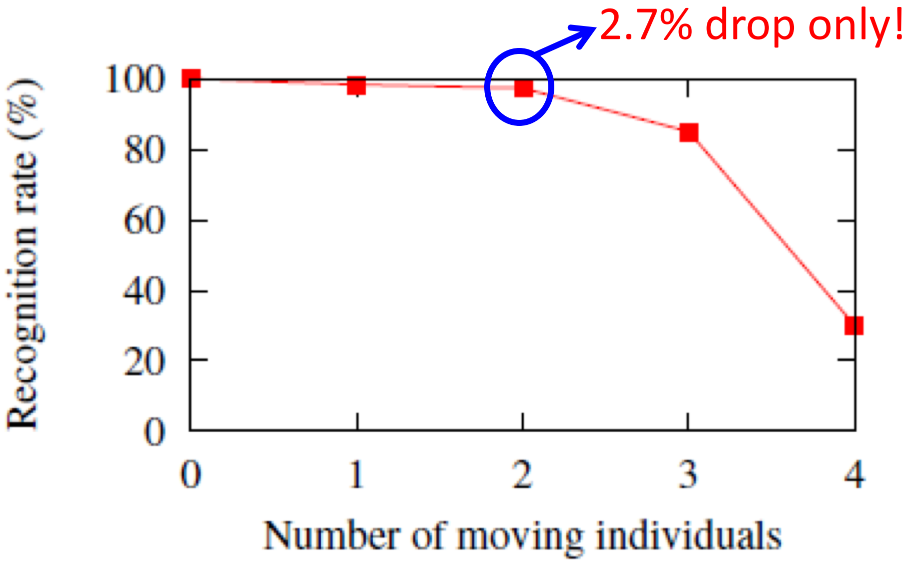
# Comparisons with Baselines

The average classification accuracy.

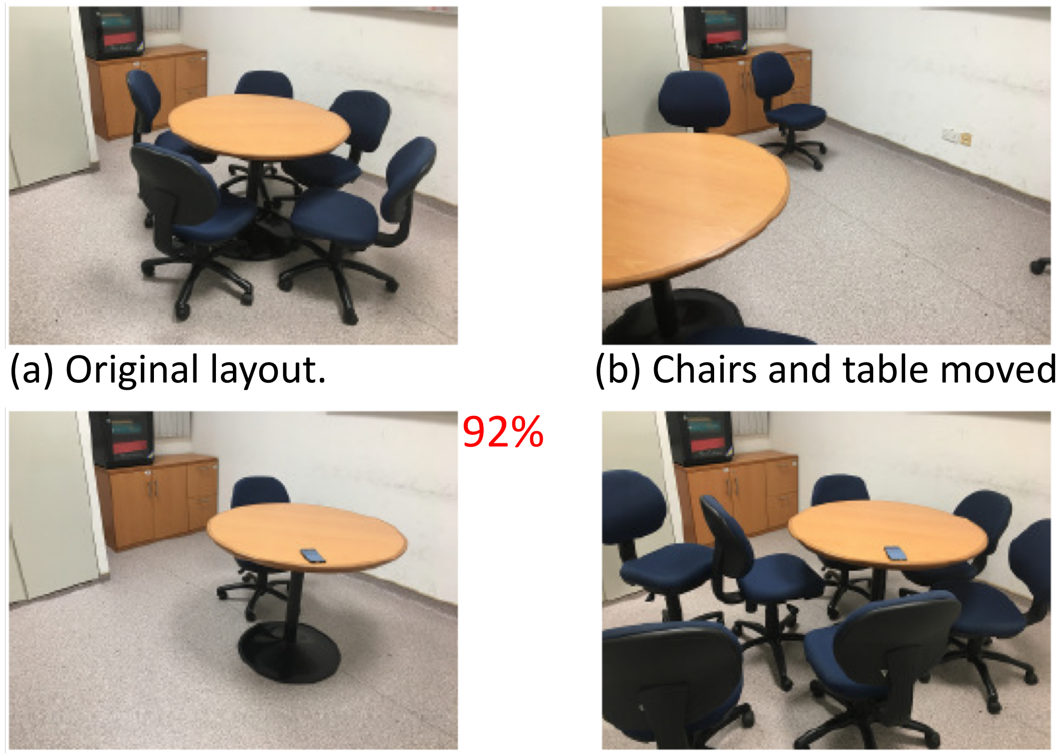
| Approach | Probe signals | Features/formats | Learning model | No music | Music |
|---|---|---|---|---|---|
| RoomSense [AH'13] | Full spectrum | Full spectrum | SVM | 76% | 39% |
| | Single tone | Full spectrum | | 83% | 27% |
| | Single tone | Narrowband | | 69% | 50% |
| Our approach | Single tone | Narrowband | CNN | 100% | 81% |

Deep learning improves the recognition accuracy even when the probe signal is very simple and the audio recording is limited to a very narrow band.

# Impact of Changes in Rooms




Recognition rate vs. the number of moving individuals in L3 ($7m^2$).




(a) Original layout.

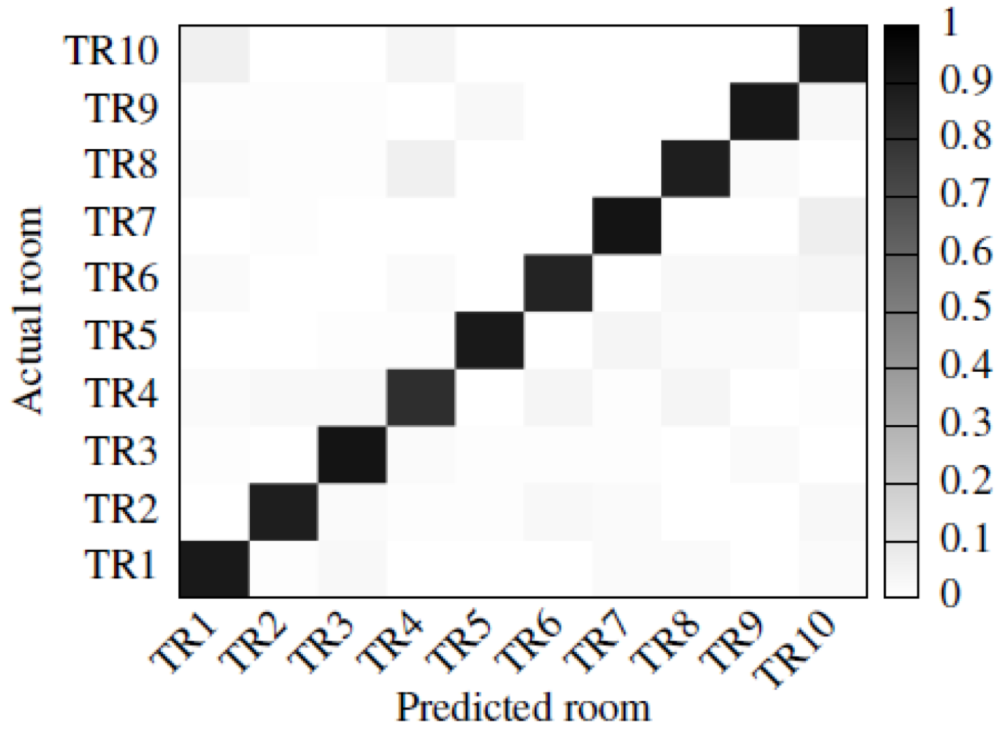
(b) Chairs and table moved.

(c) Chairs removed.

(d) More chairs added.

Furniture changes in L3.

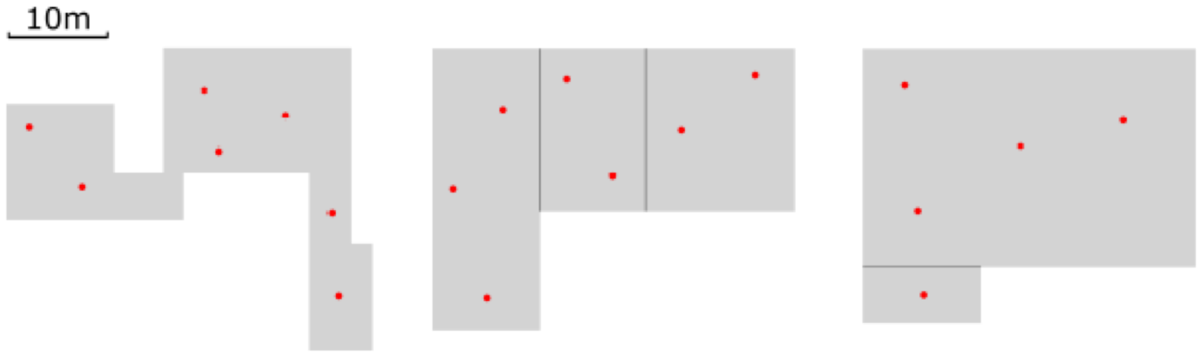
# Evaluation in Similar Rooms



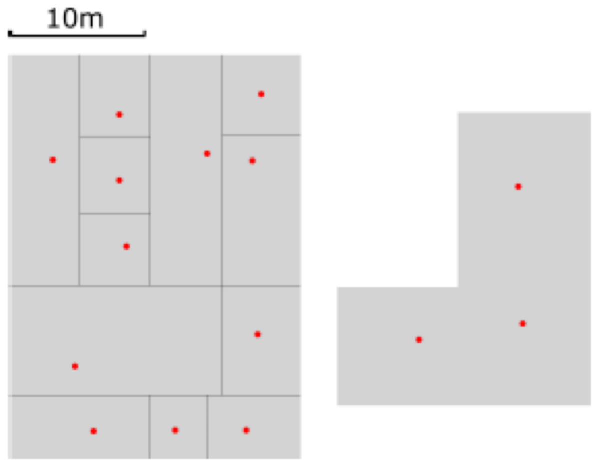

Confusion matrix of our approach in recognizing 10 similar teaching rooms (TR).

- TR1 to TR10 have the similar size and furniture
  - Our approach achieves an average accuracy of 88.9%.

# Evaluation Results in Two Museums




Museum-A floor plan and data collection spots (red points).




Museum-B floor plan and data collection spots (red points).

- Museum-A is generally quite with few visitors walking around. The average spot recognition accuracy is 99%.
- Museum-B is crowded and has background music. The average spot recognition accuracy is 89%.

# Conclusion

- Narrowband, short-time probing and recording
- High/good accuracy
  - 99.7%: 22 residential/office rooms
  - 97.7%: 50 residential/office rooms
  - 99.0%: 19 spots in a quiet museum
  - 89.0%: 15 spots in a crowded museum
- Much improved robustness against interfering sounds