See Through Walls with Wi-Fi

•••

Fadel Adib and Dina Katabi
(Massachusetts Institute of Technology)

SIGCOMM ‘13

Date 19 May 2025

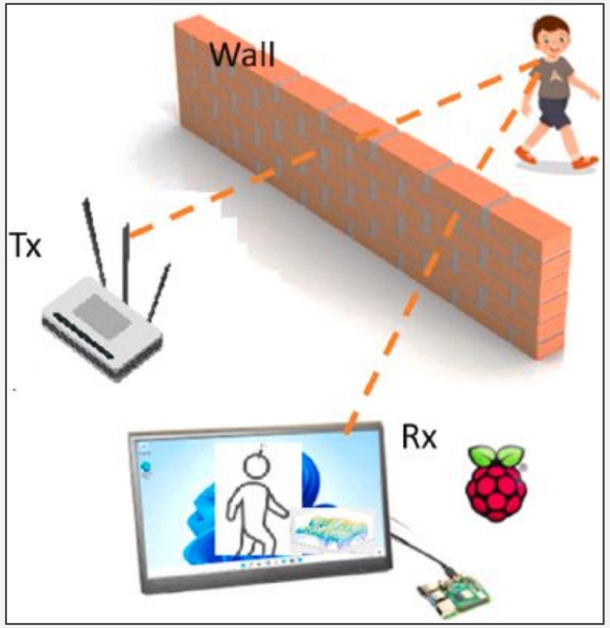
Wonseok

Table of Contents

Overview

- Background
  - Reflections of signals are imprinted with a signature of what's inside a closed room
  - Limitations of existing systems
    - An 2.4 meters <u>long antenna array</u>
      - To achieve smaller **Angular Resolution**
    - At 2 GHz <u>ultra wide bandwidth</u>
      - To eliminate **Flash Effect**

- Terms
  - Angular Resolution
    - The smallest angle that can reliably detect two objects
      - The longer antenna array, the smaller angular resolution
  - Flash Effect
    - Strong reflections from nearby surfaces
      - The wider the bandwidth, the less flash effect



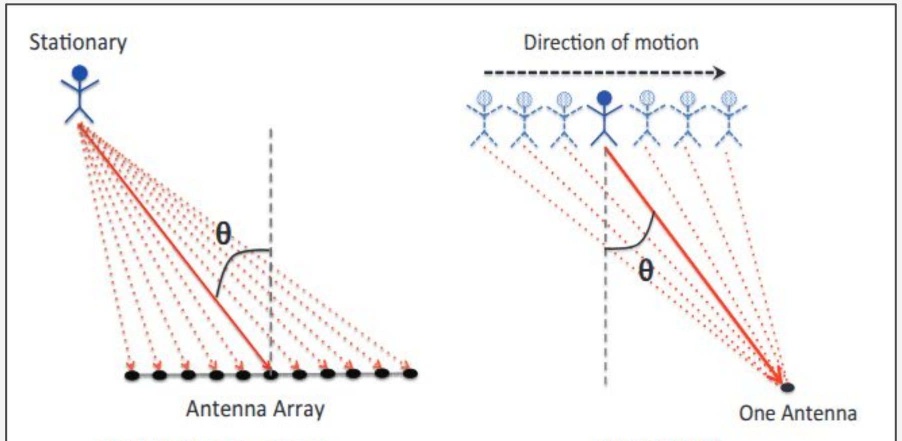
Core Problems

- Ultra-wide Bandwidth (2GHz)
  - How to eliminate flash effect without using GHz bandwidth?
  - Use MIMO → nulling to eliminate reflections (signals)
    - 1. (Tx ↔ Rx) Measure the CSI matrix
    - 2. (Tx → Rx) Null the signals
      - Reflections off static objects are nulled

    ⇢ Received signals are only reflections of moving objects

- Long Antenna Array (2.4m)
  - How to track moving objects without an antenna array?
  - Use ISAR* → virtual antenna array in snapshots
    - 1. Take snapshots of phase shift of a moving target
    - 2. Estimate a steering vector fits with θ

  ⇢ A moving target acts as an antenna array

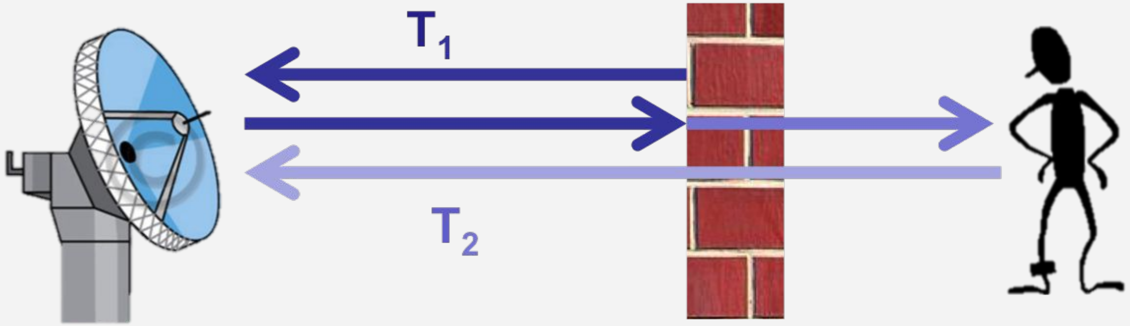


(a) Antenna Array (b) ISAR

*ISAR: Inverse Synthetic Aperture Radar

Wi-Vi

- Hardware Requirements
  - OFDM signals at 2.4GHz
  - 3-antenna MIMO (2 Tx + 1 Rx)
  - Directional antennas



- Main Components
  - 1. MIMO Nulling
    - To eliminate the **flash effect** off the wall
  - 2. Inverse SAR
    - To track a moving object and to detect gestures

Building Materials	2.4 GHz
Glass	3 dB
Solid Wood Door 1.75 inches	6 dB
Interior Hollow Wall 6 inches	9 dB
Concrete Wall 18 inches	18 dB
Reinforced Concrete	40 dB

- Challenges
  - Severe **RF attenuation** and low reflection coefficient of target object
  - **Direct signals** from Tx antennas to Rx antenna

Eliminate Flash Effect

- 1. Initial Nulling
  - a. Measure the CSI h_1, h_2 of each antenna
  - b. Pre-code only the second antenna as $\rho = -\frac{\hat{h_1}}{\hat{h_2}}$ ⟶ $$h_{receive} = h_1 + h_2(-\frac{\hat{h_1}}{\hat{h_2}}) \approx 0$$

⇢ Ideally, nulling the signals reflected from static objects

- 2. Power Boosting
  - Initial nulling also reduces the received signal power
    - Making it indistinguishable from noise

⇢ Boost the transmit power to increase SNR (roughly by 12 dB)

- 3. Iterative Nulling
  - Power boosting also amplifies residual reflections
  - Why not first boost the power and measure? → ADC saturation

⇢ Iteratively re-estimate h'_1, h'_2 with the power boosted

Iterative Nulling

- Problem
  - The receiver gets the combined CSI
    - One equation and two unknowns

$$h_{receive} = h_1 + h_2\left(-\frac{\hat{h_1}}{\hat{h_2}}\right) \approx 0$$

- Solution
  - Assume one of $\hat{h_1}$ and $\hat{h_2}$ is accurate
    - Errors of $\hat{h_1}$ and $\hat{h_2}$ are much smaller than themselves
  - Each Tx → Rx iteration, alternate $\hat{h_1}$ and $\hat{h_2}$ **in turn**

$$\hat{h_1}^{(i+1)} = h_{receive} + \hat{h}_{1(i)}$$

$$\hat{h_2}^{(i+1)} = \left(1 - \frac{h_{receive}}{\hat{h_1}^{(i)}}\right)\hat{h_2}^{(i)}$$

- Lemma
  - $\hat{h_1}$ and $\hat{h_2}$ converge exponentially fast

$$Assume\ that\ \left|\frac{\hat{h_2} - h_2}{h_2}\right| < 1, then,\ after\ i\ iterations,$$

$$|h_{res}^{(i)}| = |h_{res}^{(0)}|\left|\frac{\hat{h_2} - h_2}{h_2}\right|^i$$

Static Object

Algorithm

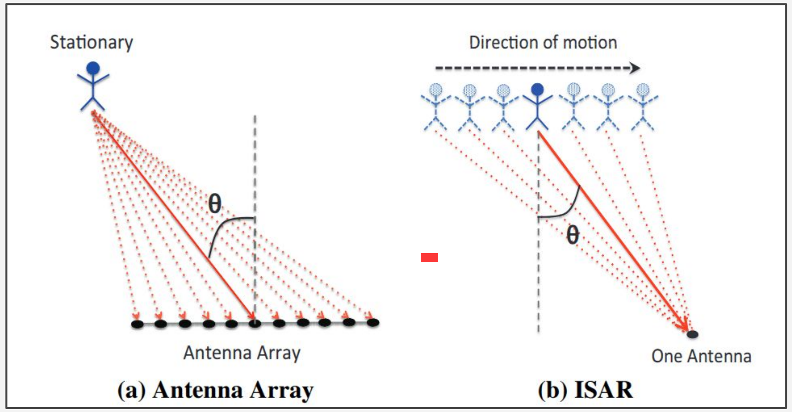
Algorithm 1 Pseudocode for Wi-Vi's Nulling

INITIAL NULLING:
▷ Channel Estimation
Tx ant. 1 sends x; Rx receives y; $\hat{h_1} \leftarrow y/x$
Tx ant. 2 sends x; Rx receives y; $\hat{h_2} \leftarrow y/x$
▷ Pre-coding: $p \leftarrow -\hat{h_1}/\hat{h_2}$
POWER BOOSTING:
Tx antennas boost power
Tx ant. 1 transmits x, Tx ant. 2 transmits px concurrently
ITERATIVE NULLING:
$i \leftarrow 0$
repeat
    Rx receives y; $h_{res} \leftarrow y/x$
    if i even **then**
        $\hat{h_1} \leftarrow h_{res} + \hat{h_1}$
    else
        $\hat{h_2} \leftarrow \left(1 - \frac{h_{res}}{\hat{h_1}}\right)\hat{h_2}$
    $p \leftarrow -\hat{h_1}/\hat{h_2}$
    Tx antennas transmit concurrently
    $i \leftarrow i + 1$
until Converges

Moving Object

Human Tracking

- Idea
  - A moving human can be treated as an antenna array
    - Successive time samples as spatial samples



- Phase Shift $\Phi(\theta, n)$
  - θ : Spatial angle
    - The angle between human to Wi-Vi and the normal to the motion
  - $\triangle$: Spacing between successive antennas
    - $\triangle = vt(v \approx 1m/s)$

$$\Phi(\theta, n) = \frac{2\pi}{\lambda} n \triangle \sin\theta$$

- Beamforming
  - The virtual antenna array performs ‘beamforming’
    - w : Number of time samples
    - $h[n]$: Channel gain

$$A[\theta, n] = \sum_{i=1}^{w} h[n+i] e^{j\Phi(\theta, n)}$$

*ISAR: Inverse Synthetic Aperture Radar

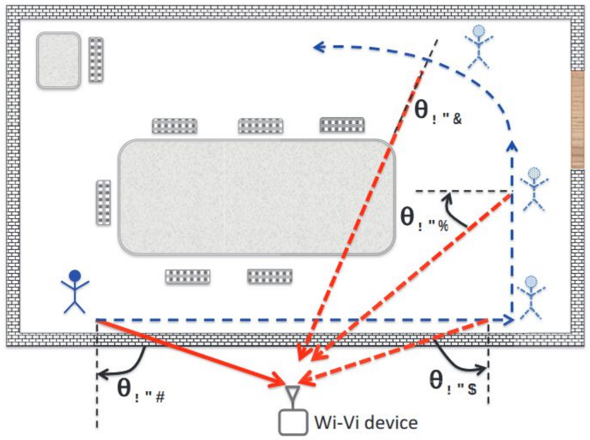
Moving Object

Beamforming

- Motion Direction
  - Estimate θ that maximizes the channel sum
    - Successive time samples as spatial samples

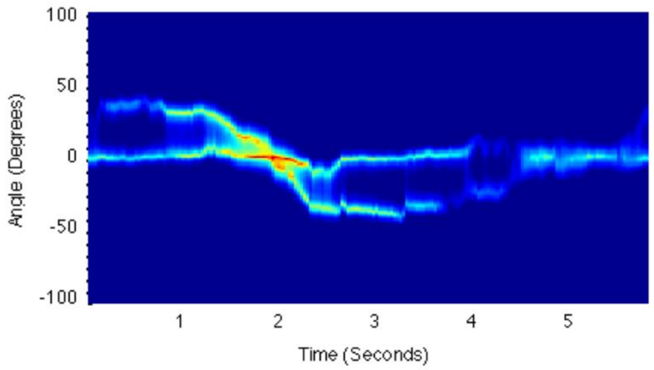
$$A[\theta, n] = \sum_{i=1}^{w} h[n+i] e^{j\Phi(\theta,n)}$$

- Experiment Setup
  - **2s**: The person crosses the Wi-Vi device
  - **2s~3s**: The person is moving away from the device
  - **3s**: The person turns inward, but signal gets weaker



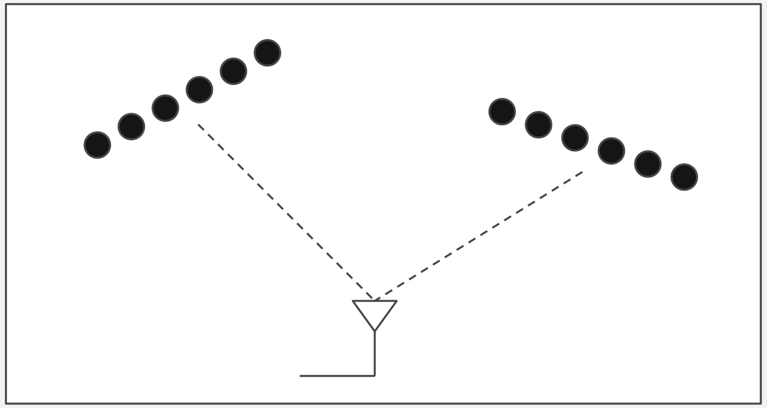
(a) **Experimental Setup**

- Heatmap Output of $A[\theta, n]$
  - Two lines are present
    - Zero line: Amplified residual reflection on power boosting
    - Curved line: Human motion



Multiple Human Tracking

- MUSIC Algorithm
  - Estimate incoming AoAs from entangled channel gain
    - Orthogonality between noise space and steering vectors
  - How come?
    - Only a steering vector matters, not distance
    - Antenna spacing $\triangle$ are assumed the same ($\triangle = vt(v \approx 1m/s)$)
      - ⇢ Limitation of this paper



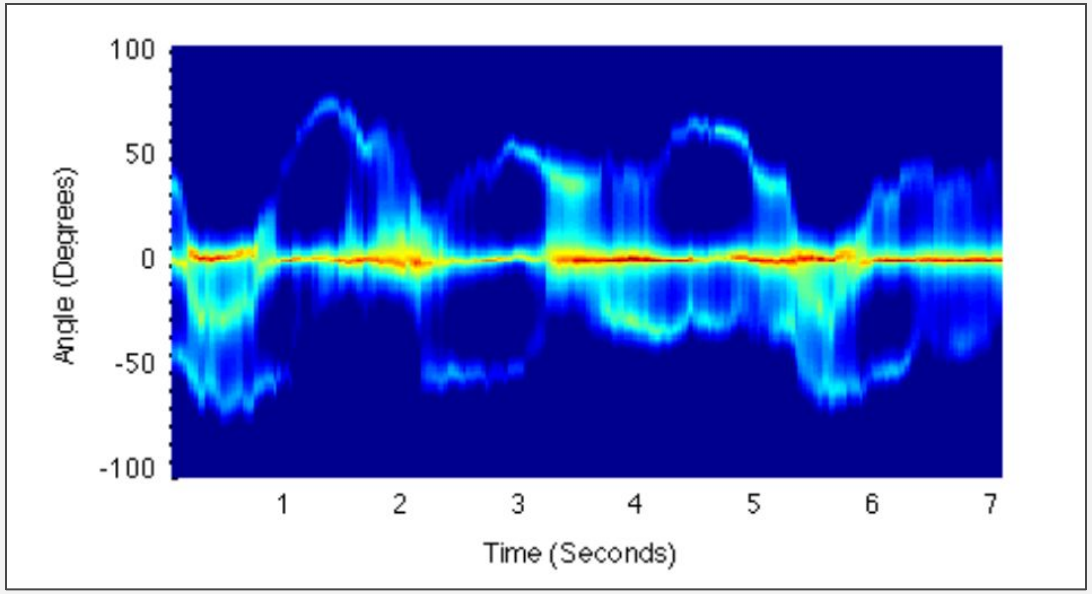
- Steps
  - 1. Create a correlation matrix $R_{w\times w} = E[hh^H]$
  - 2. Find noise space from eigen-decomposition
  - 3. Project angles onto the noise space
  - 4. Find peaks at θ
    - Orthogonality at θ

$$A'[\theta, n] = \frac{1}{\sum_{k=1}^{K} || \sum_{i=1}^{w} e^{-j\frac{2\pi}{\lambda} i\Delta \sin\theta} U_N[n](i,k)||^2}.$$

$$P_{MUSIC}(\theta) = A'[\theta, n] = \frac{1}{||U_n^H \vec{a}_n(\theta)||}$$

Number of Humans

- Number of Humans
  - Heuristically found
    - N people → N lines on the graph
      - Try machine learning
  - Related to VAR[θ] in statistical sense



- Machine Learning (Classification)
  - How many people given VAR[θ]?
    - **Input**: time-averaged angular variance

$$x = \frac{1}{T}\sum_{n=1}^{T} VAR_n[\theta]$$

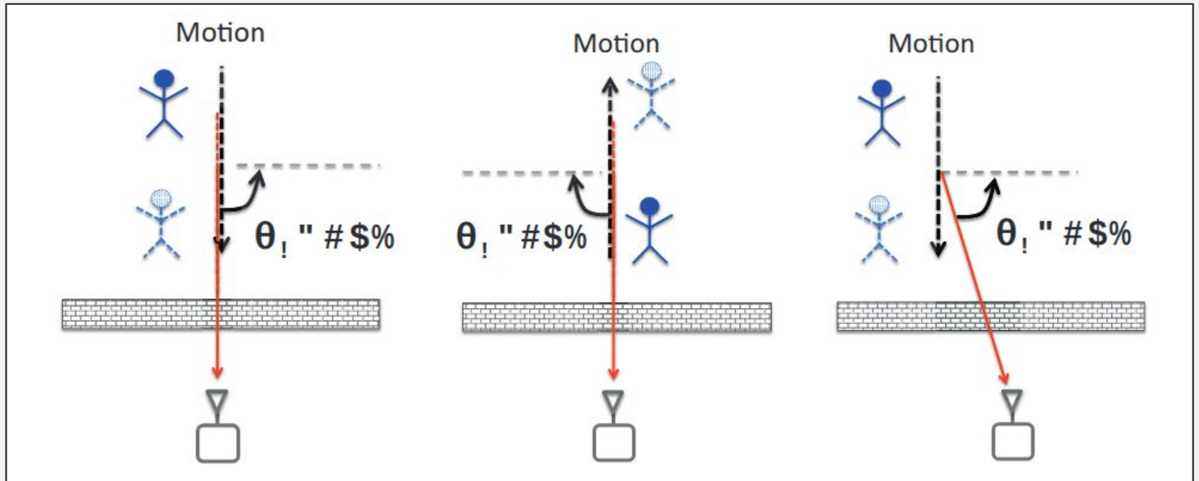
    - **Output**: 0, 1, 2, or 3 humans

$$E[n] \approx C[n] = \sum_{\theta=-90}^{90} \theta \cdot 20 \log_{10} A'[\theta, n]$$

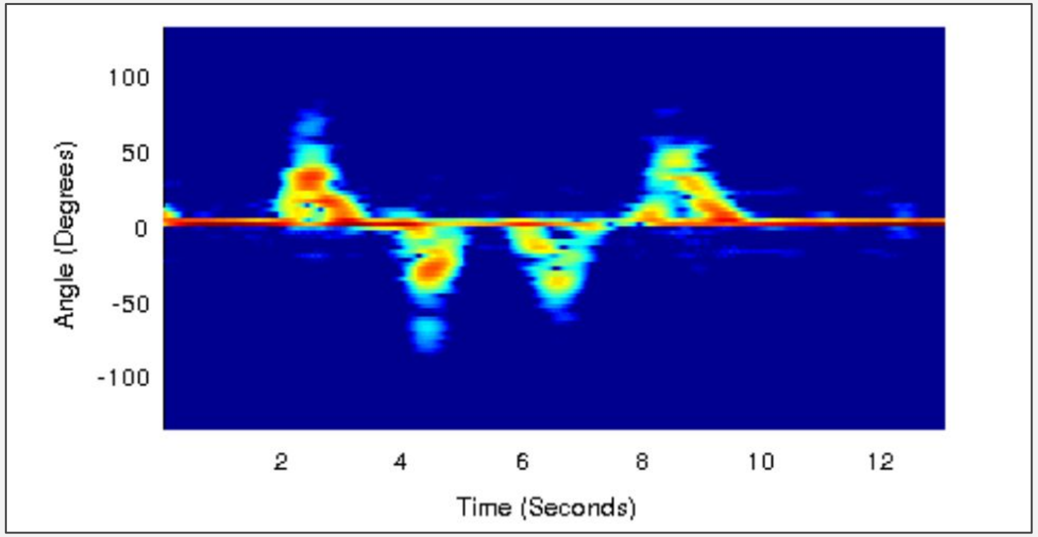
$$VAR_n[\theta] = E[\theta^2] - E^2[\theta]$$

Gesture Encoding

- Gestures
  - ‘0’ : A step forward + a step backward
    - MUSIC peaks at positive θ first
  - ‘1’: A step backward + a step forward
    - MUSIC peaks at negative θ first



(a) Forward (b) Backward (c) Slanted



Gesture Decoding

- Decoding Steps
  - 1. Apply ‘matched filter’ on the heatmap $A[\theta, n]$
    - Matched filter’s template: ▲▼

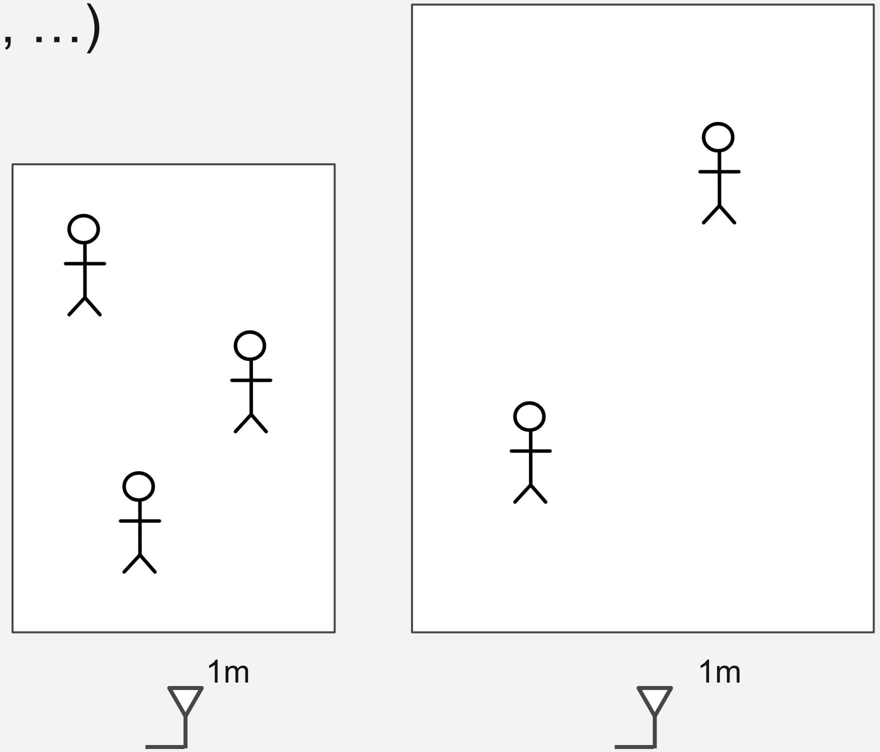
$$\boxed{Matched\ Output = \sum_{\theta=90}^{90} A[\theta, n] * template[\theta]}$$

  - 2. Apply peak detector on the output

Experiment

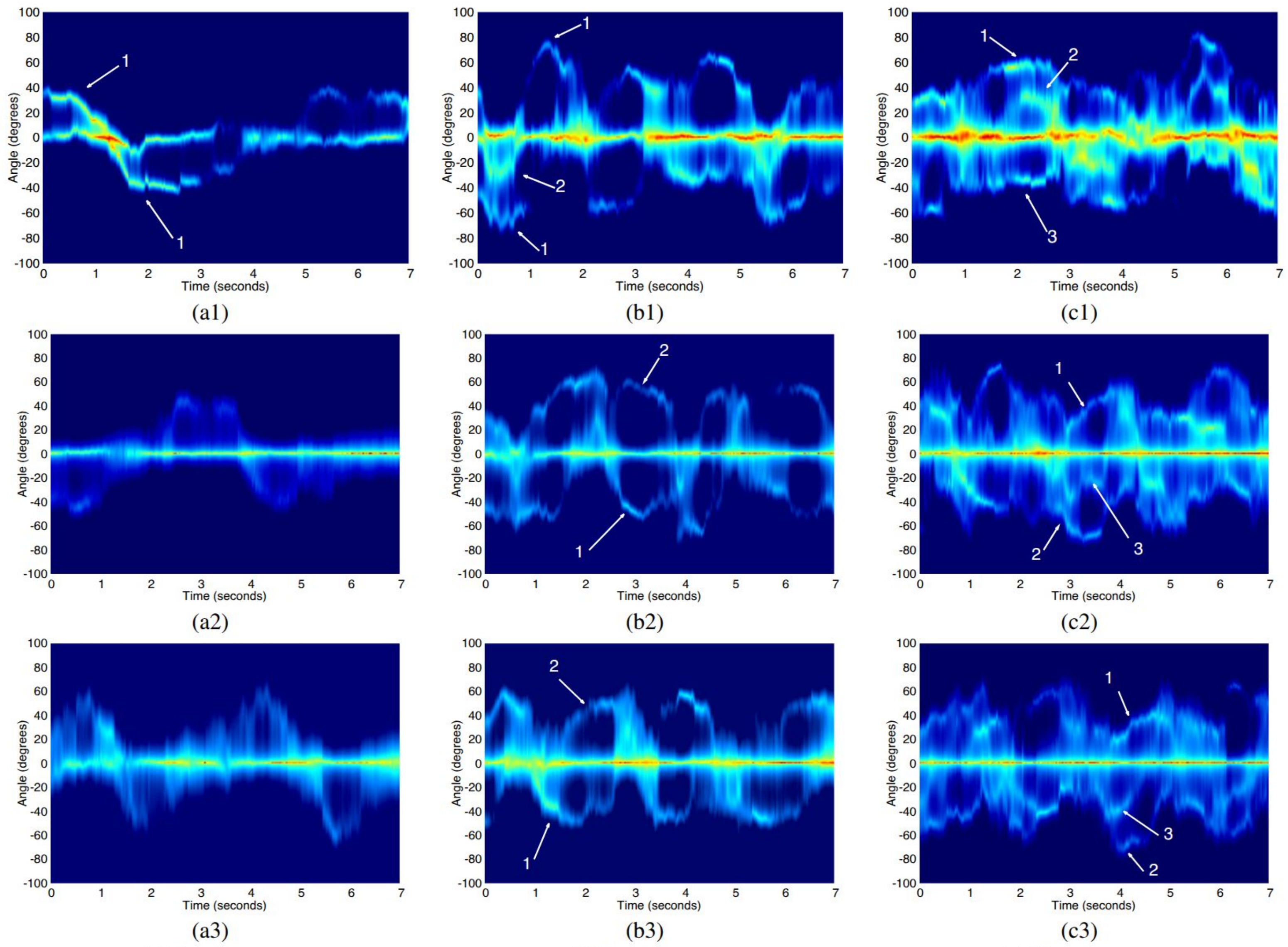
Environment

- Locations
  - Two conference rooms with standard furniture (tables, chairs, …)
    - Room A: 7 X 4 meters
    - Room B: 11 X 7 meters
    - Wall: 6-inch, hollow
  - Wi-Vi device 1m away from a wall



Experiment

Tracking



(a) One Human (b) Two Humans (c) Three Humans

- Brightness
  - ⇢ Indicates distance
- Fuzziness
  - ⇢ Worsen by body partsz

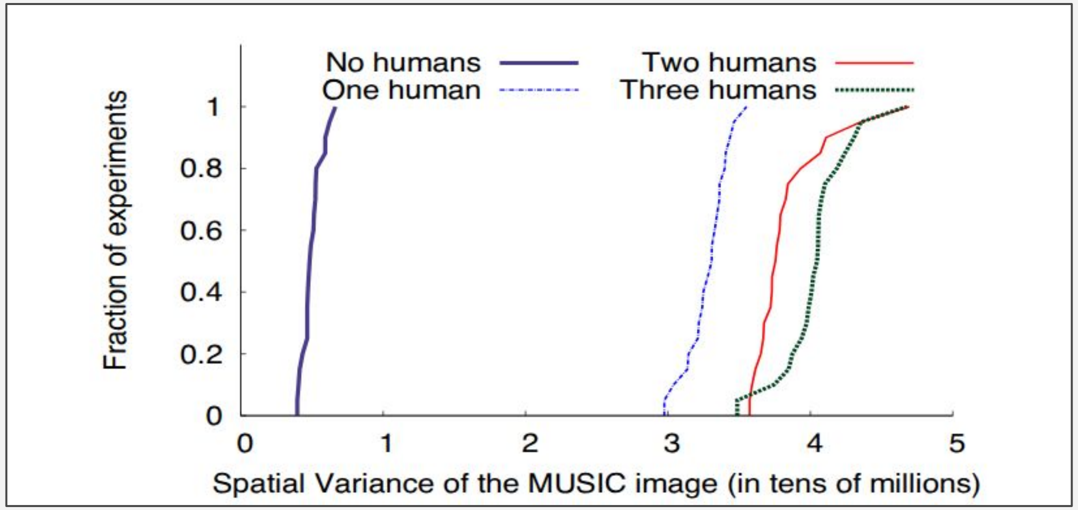
$$P_{MUSIC}(\theta) = A'[\theta, n] = \frac{1}{||U_n^H \vec{a}_n(\theta)||}$$

Automatic Detection

- CDF (20 Trials each)
  - **x-axis**: spatial variance $\frac{1}{T}\sum_{n=1}^{T} VAR_n[\theta]$ ⟶
  - **y-axis**: fraction of experiments

  ⇢ **y** trials had variance ≤ **x**

$$VAR[n] = \sum_{\theta=-90}^{90} \theta^2 C[n] - C[n]^2$$



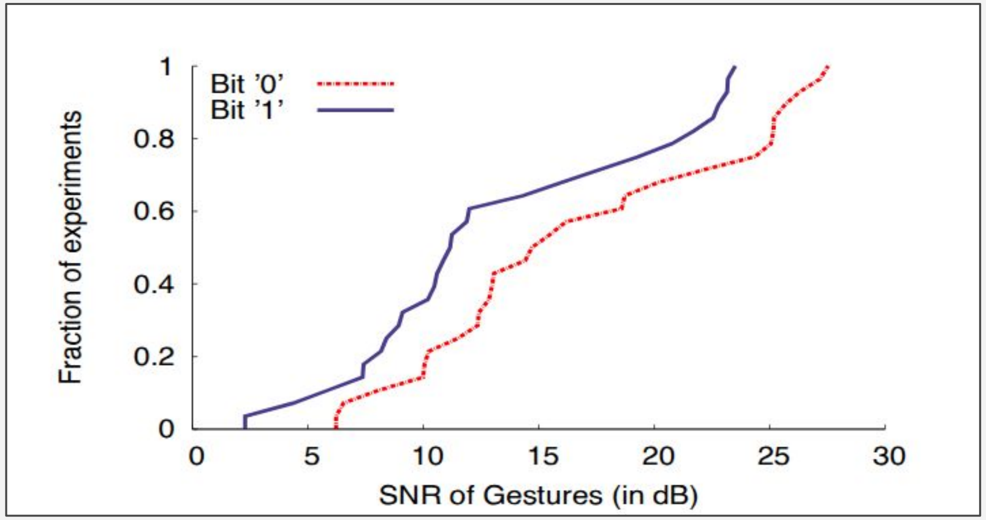
- Observations
  - Spatial variance is higher with more moving bodies
  - Steep curves → consistent variance
  - The separation between successive CDFs ▼

- Accuracy
  - Training set on Room A, testing set on Room B

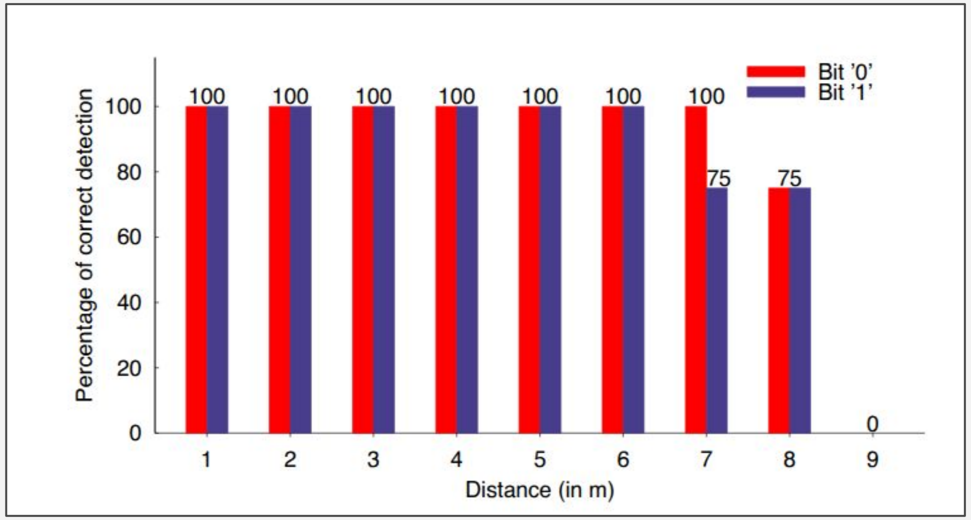
Actual \ Detected	0	1	2	3
0	100%	0%	0%	0%
1	0%	100%	0%	0%
2	0%	0%	85%	15%
3	0%	0%	10%	90%

Gesture Decoding

- CDF
  - ‘0’ bit’s SNR is higher than ‘1’
    - Taking a step forward first -> average closer
    - Taking a step backward is naturally harder
      - People tend to take smaller steps



- Accuracy
  - 9m: 0% -> not enough energy to detect from the noise
    - 20mW used (USRP)
  - Never mistook ‘0’ or ‘1’ bit



Conclusion

- Impact
  - First low-bandwidth through-wall radar
    - Military purpose and ultra wide-band so far
  - MIMO interference nulling
    - Cancelling “flash effect”

- Limitation
  - Moving objects have constant speed
  - Only detect the direction of motion
    - Cannot detect if a person stops
    - Cannot estimate distance