香港科技大學
THE HONG KONG UNIVERSITY OF SCIENCE AND TECHNOLOGY

THE DEPARTMENT OF COMPUTER SCIENCE & ENGINEERING
計算機科學及工程學系

We Can Hear You with Wi-Fi !

Guanhua Wang

Yongpan Zou, Zimu Zhou, Kaishun Wu, Lionel M. Ni

Hong Kong University of Science and Technology

Advanced Research in ISM band

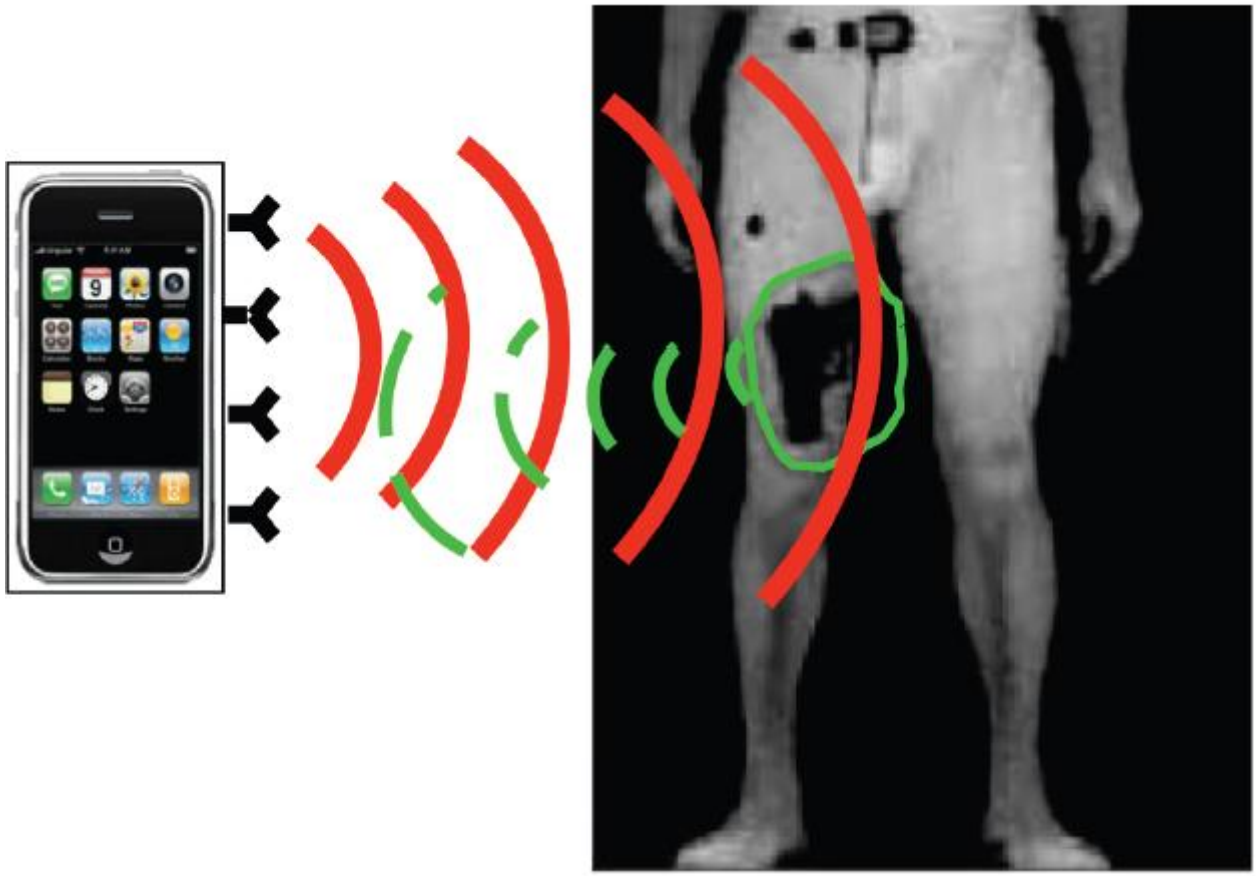
- Localization

Advanced Research in ISM band

- Gesture Recognition

Advanced Research in ISM band

- Object Detection



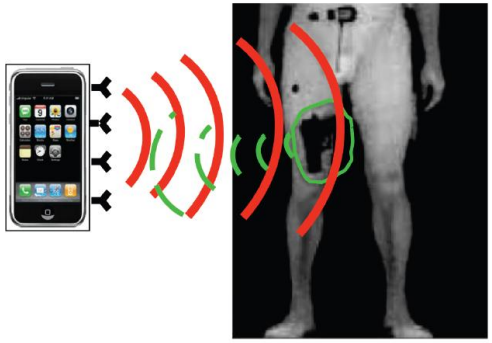
Full Duplex Backscatter. In hotnets 2013

Advanced Research in ISM band

- Localization

- Gesture recognition

- Object Classification



They enable Wi-Fi to “SEE” target objects.

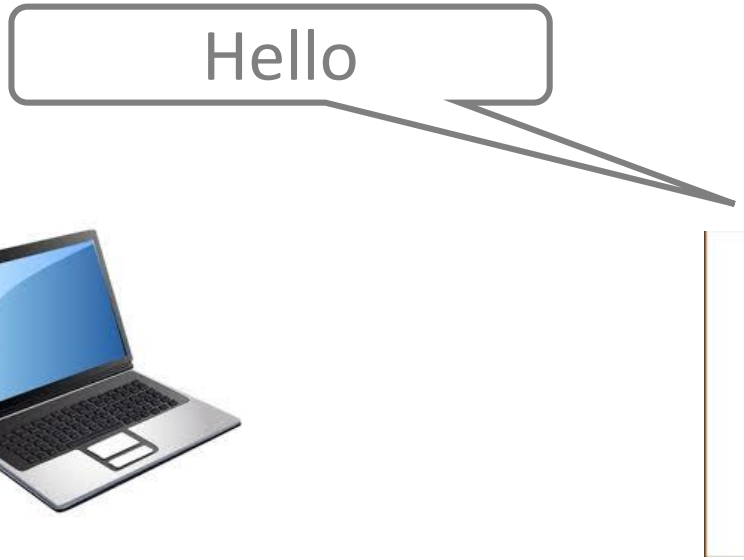
Can we enable Wi-Fi signals to HEAR talks?

Can we enable Wi-Fi signals to HEAR talks?

WiHear

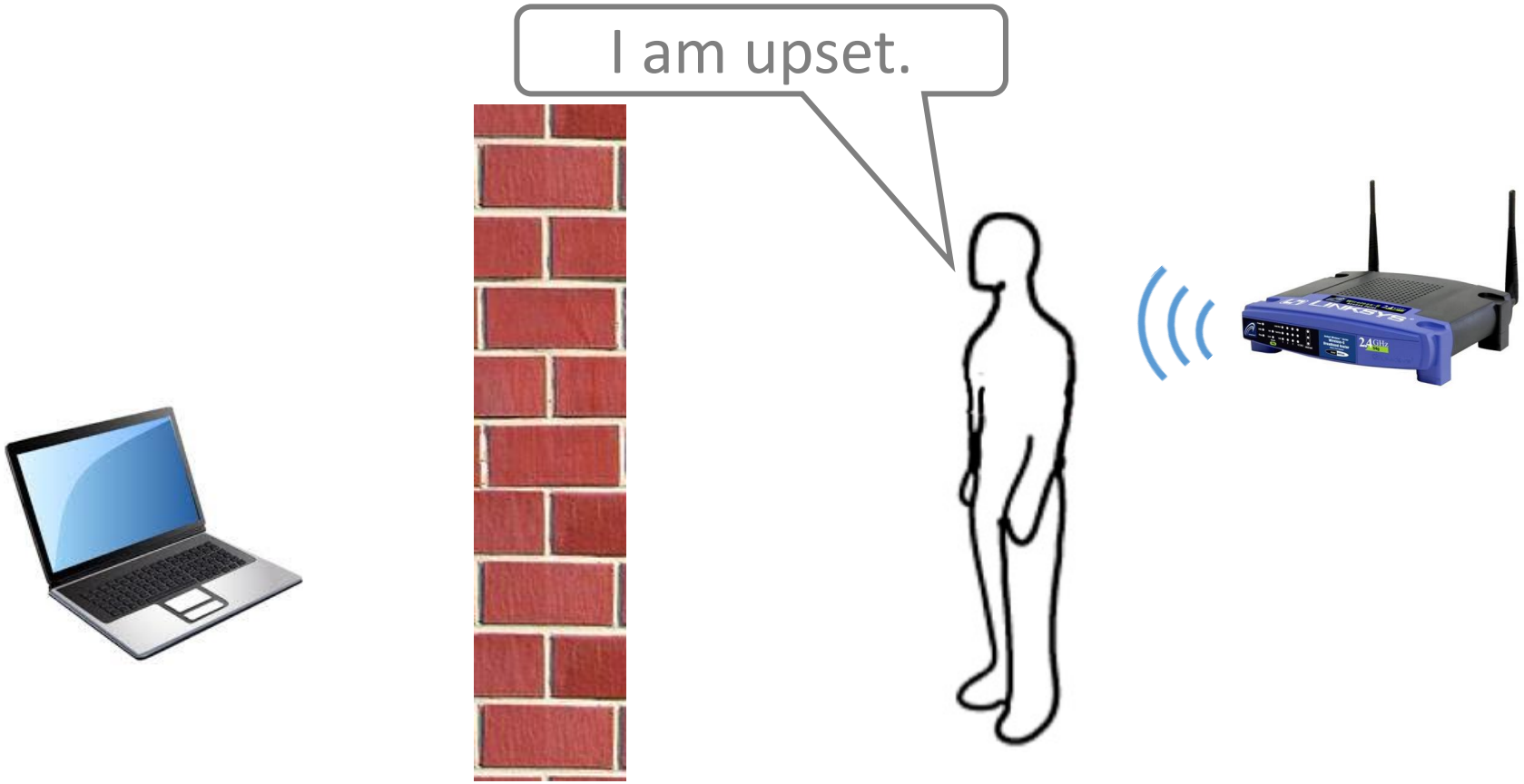
What is WiHear?

“Hearing” human talks with Wi-Fi signals



Non-invasive and device-free

Hearing through walls and doors



Understanding complicated human behavior (e.g. mood)

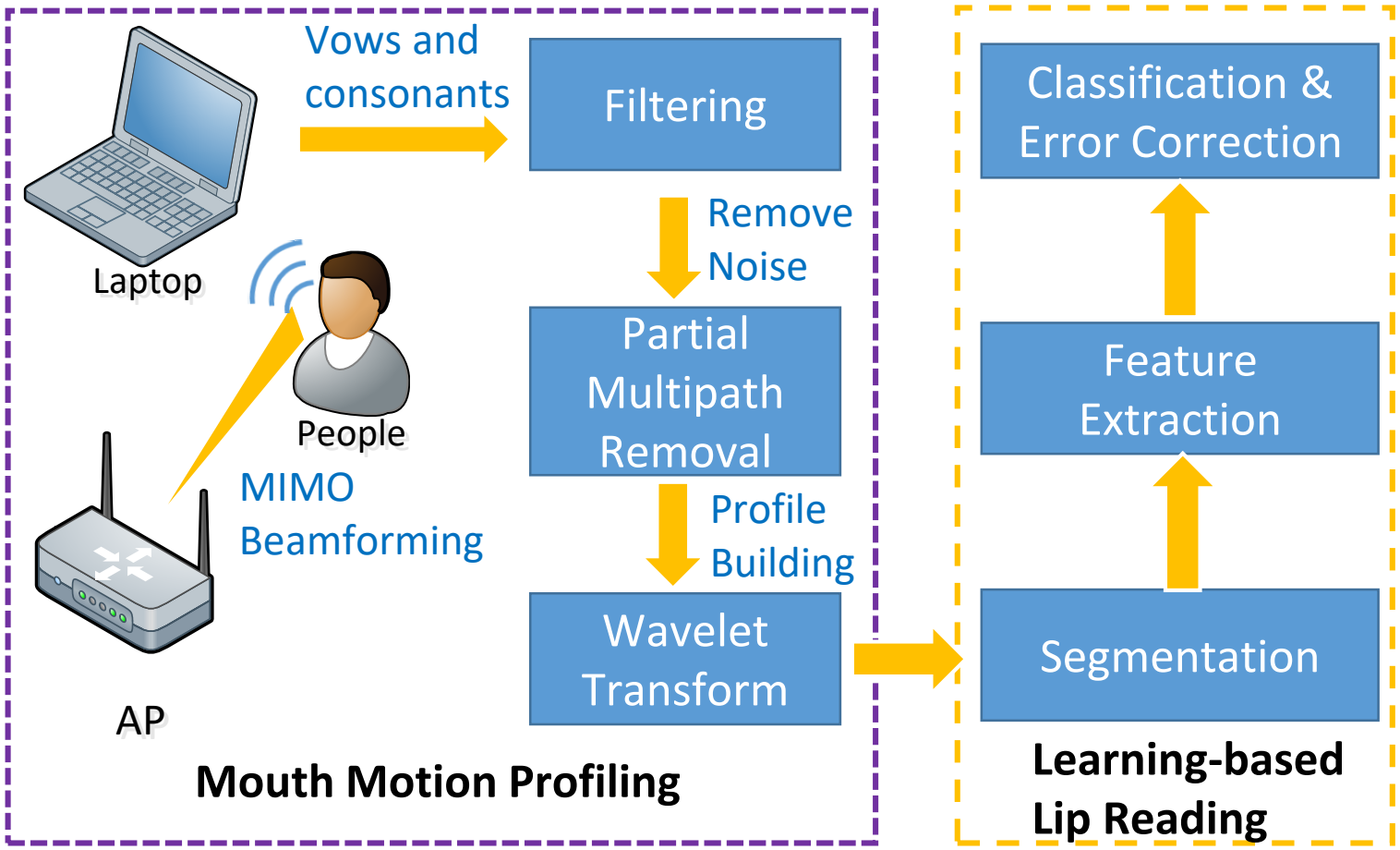
Hearing multiple people simultaneously

MIMO Technology

Easy to be implemented in commercial Wi-Fi products

How does WiHear work?

WiHear Framework



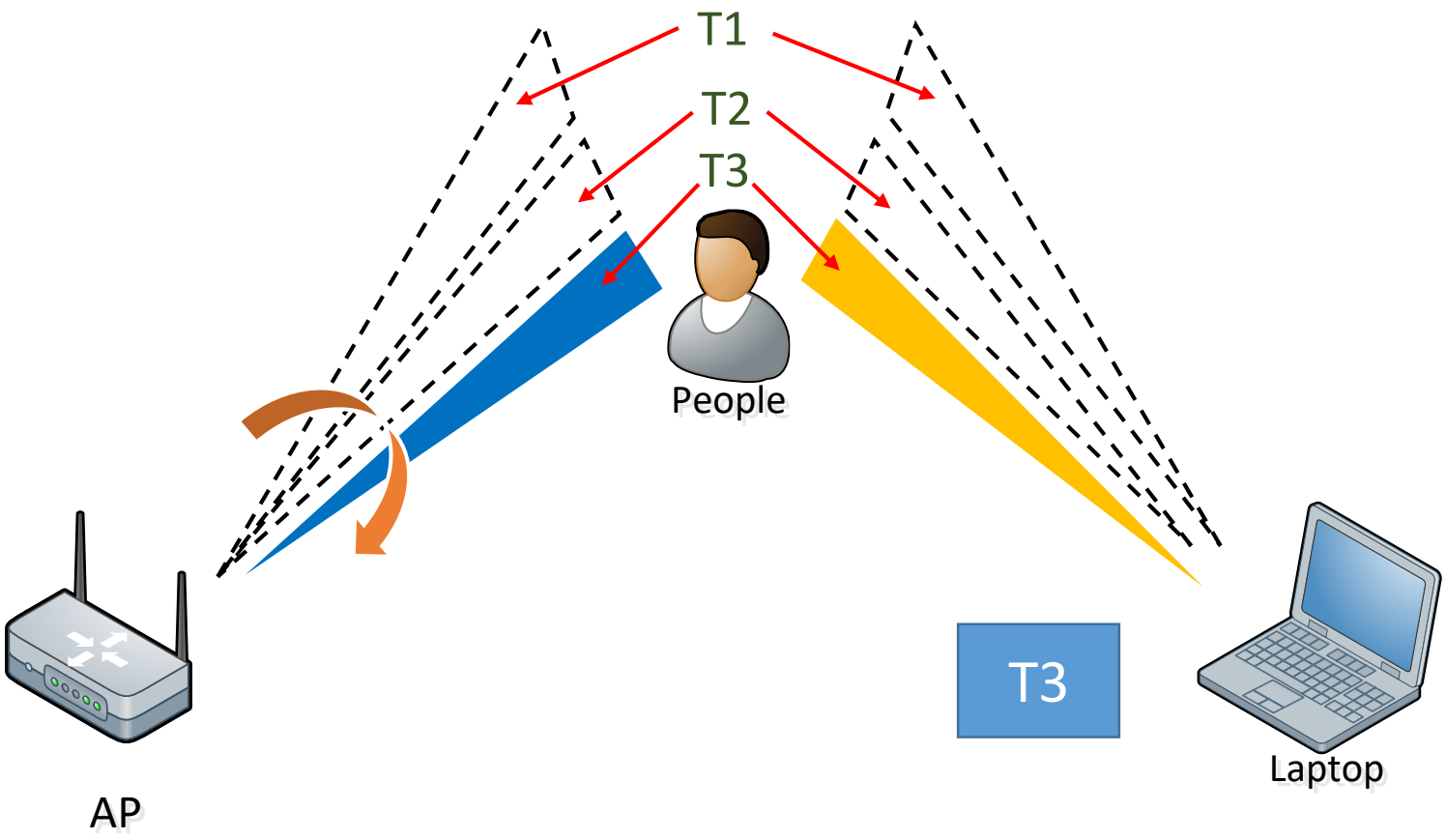
Mouth Motion Profiling

- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Mouth Motion Profiling

- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Locating on Mouth



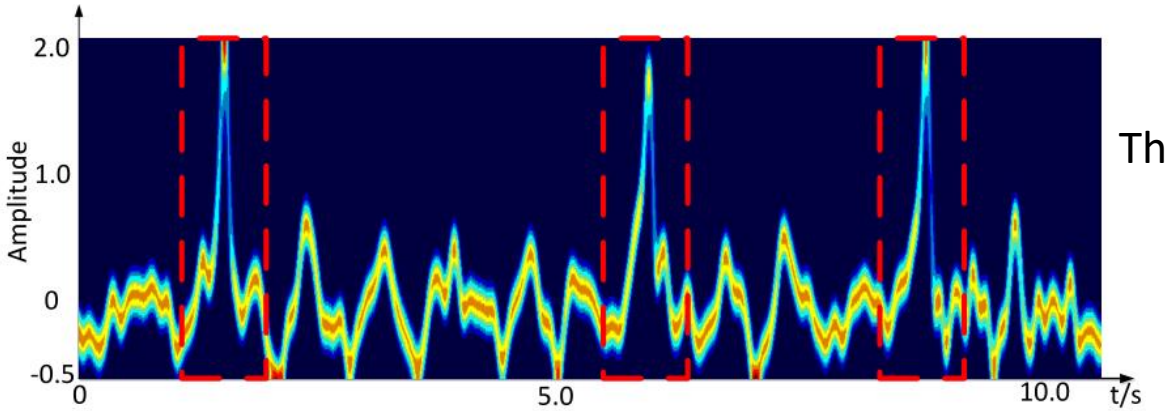
Mouth Motion Profiling

- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Filtering Out-Band Interference

- Signal changes caused by mouth motion: 2-5 Hz

- Adopt a 3-order Butterworth IIR band-pass filter
  - ➢Cancel the DC component
  - ➢Cancel wink issue (<1 Hz)



The impact of wink (as denoted in the dashed red box).

  - ➢Cancel high frequency interference

Mouth Motion Profiling

- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Partial Multipath Removal

- Mouth movement: Non-rigid
- Covert CSI (Channel State Information) from frequency domain to time domain via IFFT
- Multipath removal threshold: >500 ns
- Convert processed CSI (with multipath < 500ns) back to frequency domain via FFT

The multipath threshold value can be adjusted to achieve better performance

Mouth Motion Profiling

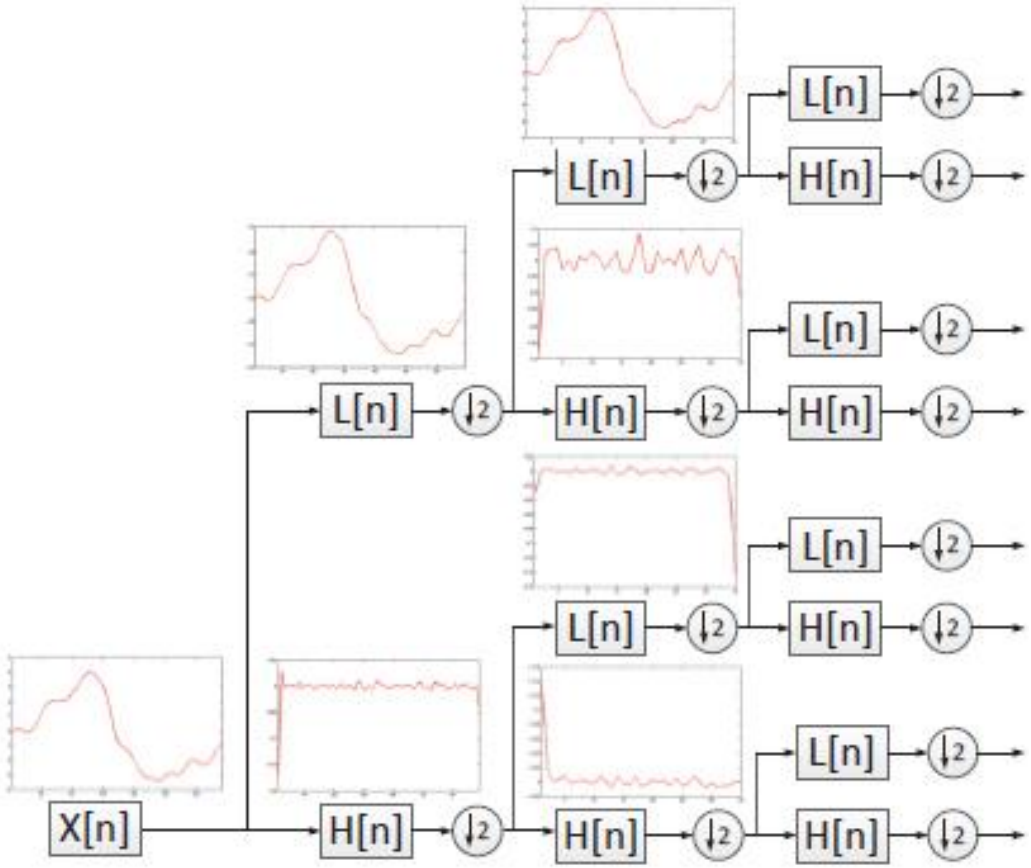
- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Mouth Motion Profiling

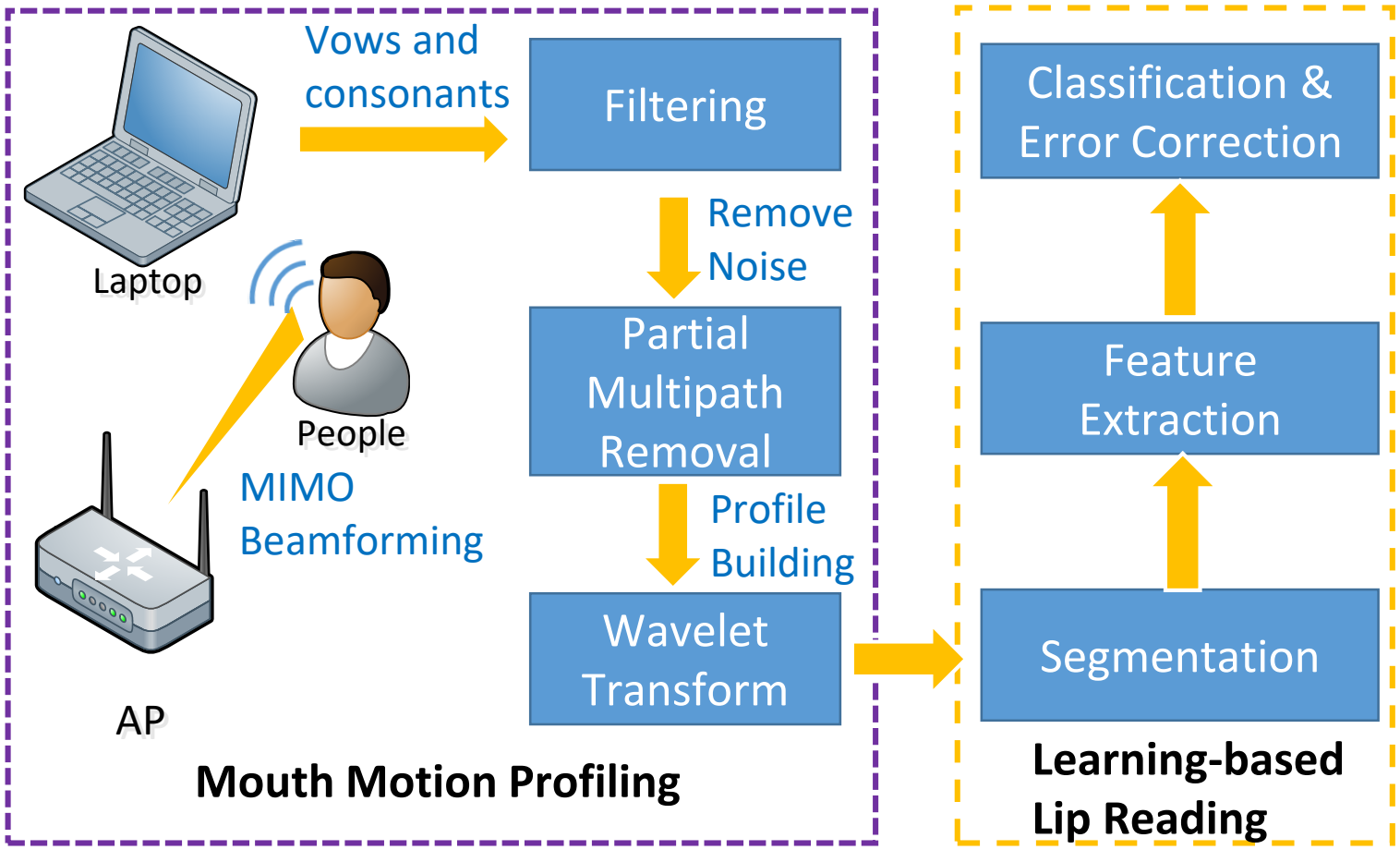
- Locating on Mouth
- Filtering Out-Band Interference
- Partial Multipath Removal
- Mouth Motion Profile Construction
- Discrete Wavelet Packet Decomposition

Discrete Wavelet Packet Decomposition

- A Symlet wavelet filter of order 4 is selected



WiHear Framework



Lip Reading

- Segmentation
- Feature Extraction
- Classification
- Context-based Error Correction

Lip Reading

- Segmentation
- Feature Extraction
- Classification
- Context-based Error Correction

Segmentation

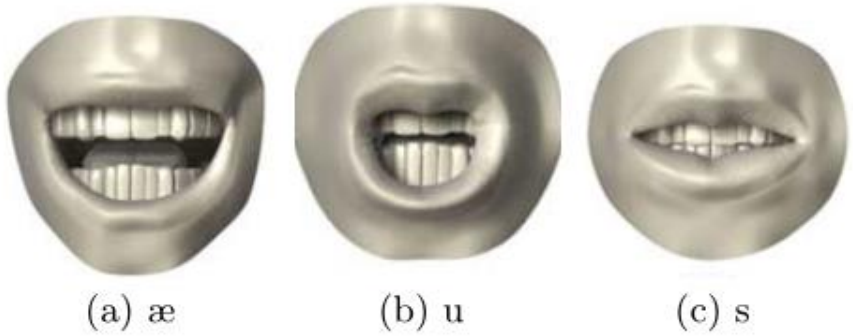
- Inter word segmentation
  - ➢ Silent interval between words
- Inner word segmentation
  - ➢ Words are divided into phonetic events

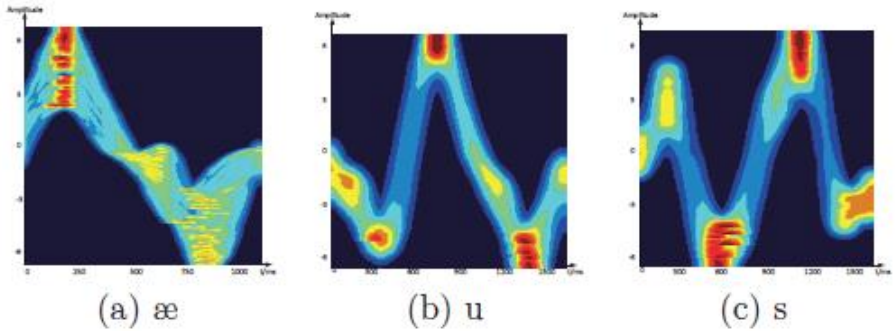
Lip Reading

- Segmentation
- Feature Extraction
- Classification
- Context-based Error Correction

Feature Extraction

- Multi-Cluster/Class Feature Selection (MCFS) scheme





Lip Reading

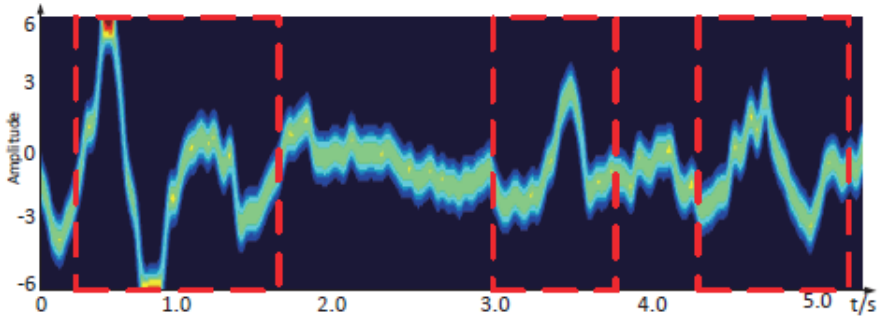
- Segmentation
- Feature Extraction
- Classification
- Context-based Error Correction

Lip Reading

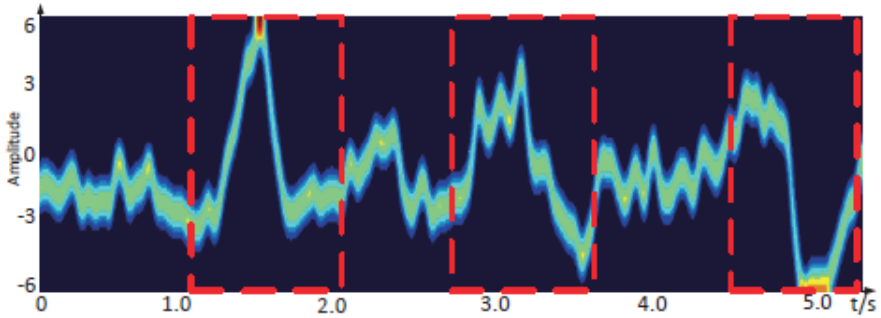
- Segmentation
- Feature Extraction
- Classification
- Context-based Error Correction

Extending To Multiple Targets

- MIMO: Spatial diversity via multiple Rx antennas
- ZigZag decoding: a single Rx antenna

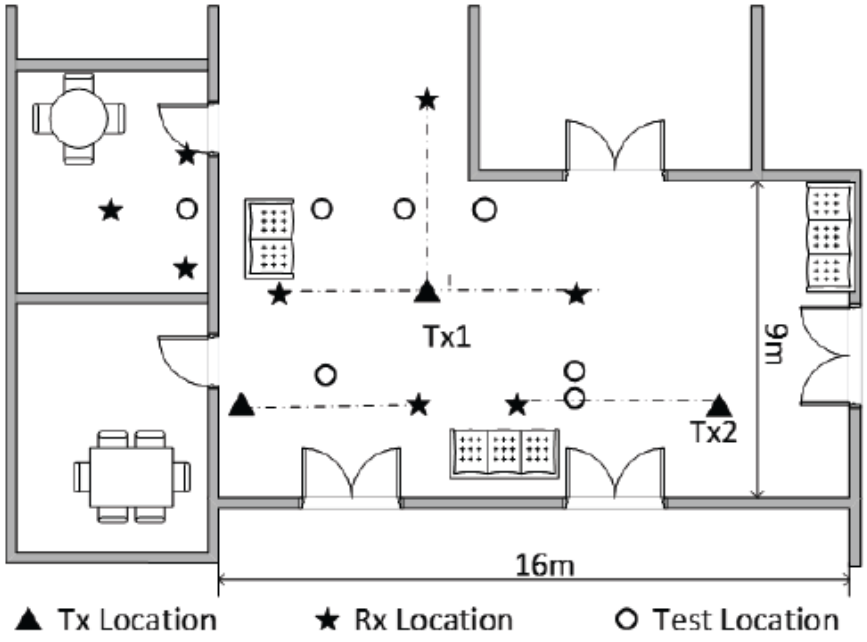


(a) Representive signals of user1 speaking

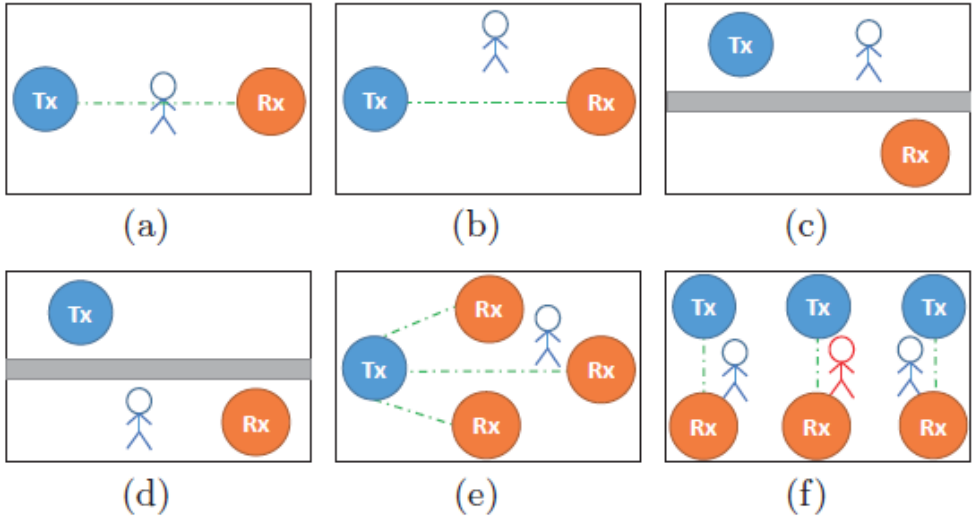


(b) Representive signals of user2 speaking

Implementation



Floor plan of the testing environment.



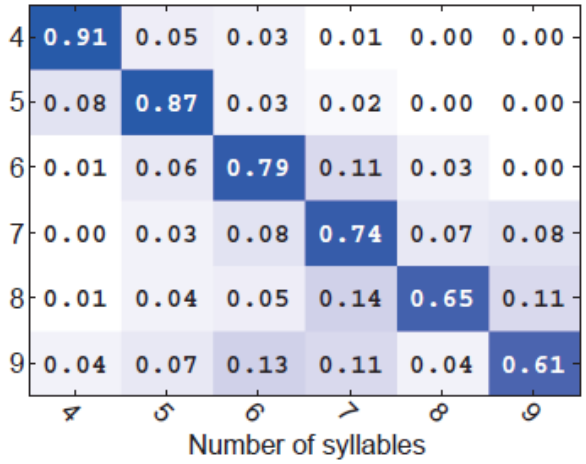
Experimental scenarios layouts. (a) line of sight; (b) non-line-of-sight; (c) through wall Tx side; (d) through wall Rx side; (e) multiple Rx; (f) multiple link pairs.

Vocabulary

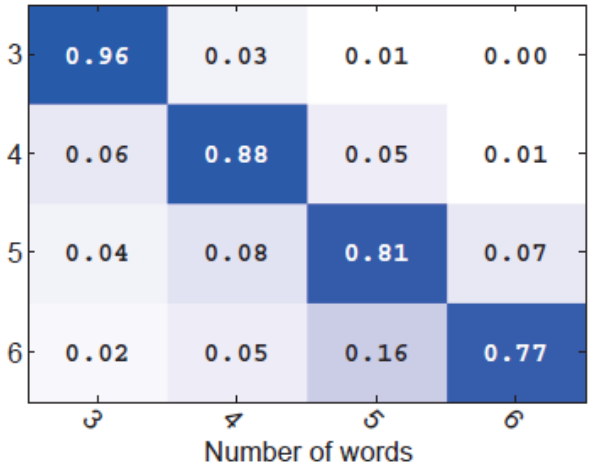
- **Syllables:**
  - [æ], [e], [i], [u], [s], [l], [m], [h], [v], [ɔ], [w], [b], [j], [ʃ].

- **Words:**
  - see, good, how, are, you, fine, look, open, is, the, door, thank, boy, any, show, dog, bird, cat, zoo, yes, meet, some, watch, horse, sing, play, dance, lady, ride, today, like, he, she.

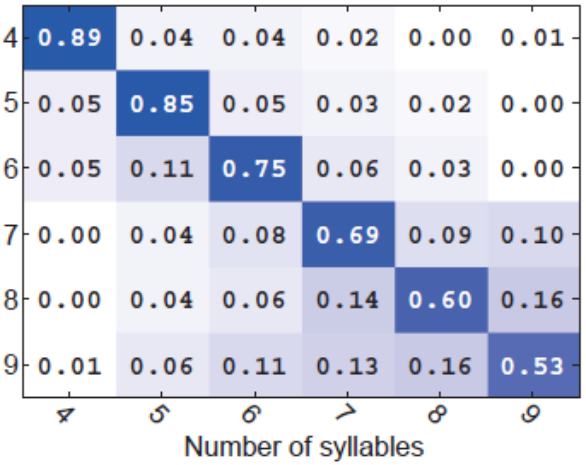
Automatic Segmentation Accuracy



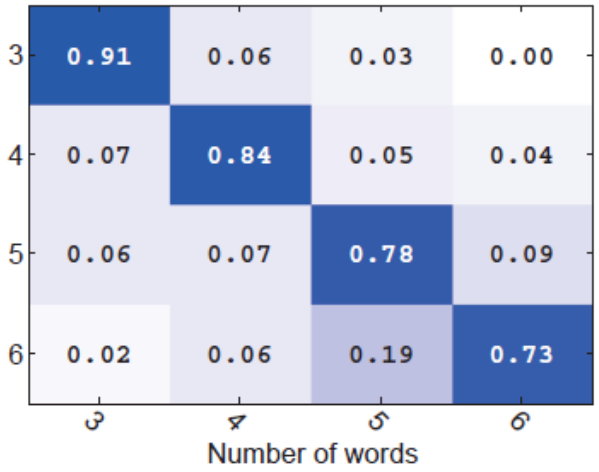
(a)



(b)



(c)



(d)

Automatic segmentation accuracy for

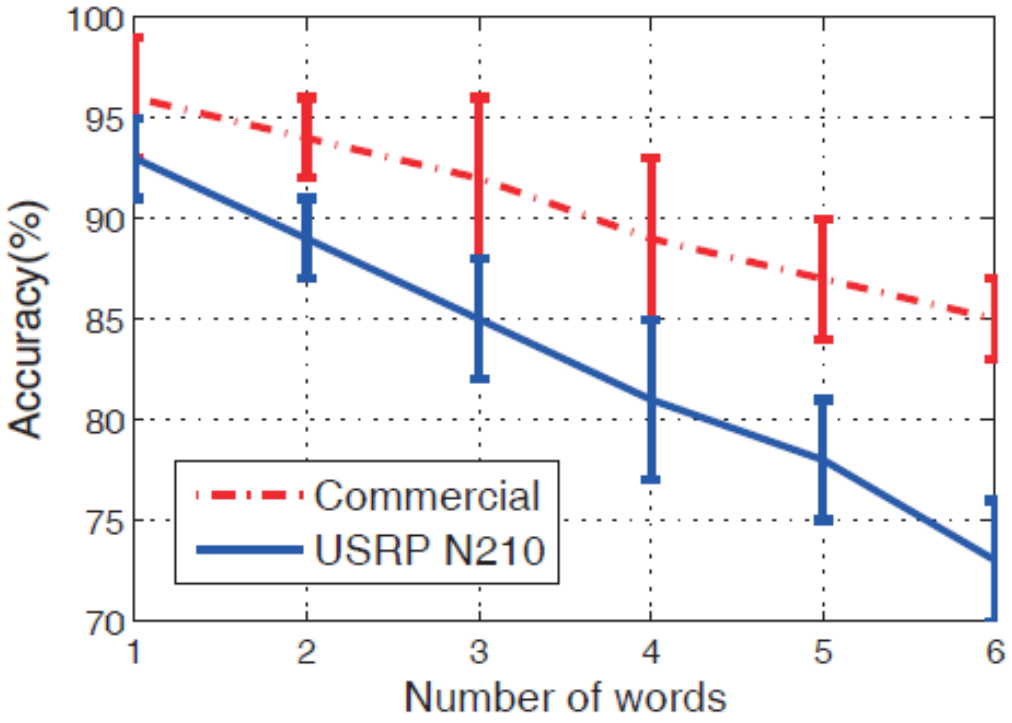
(a) Inner-word segmentation on commercial devices

(b) Inter-word segmentation on commercial devices

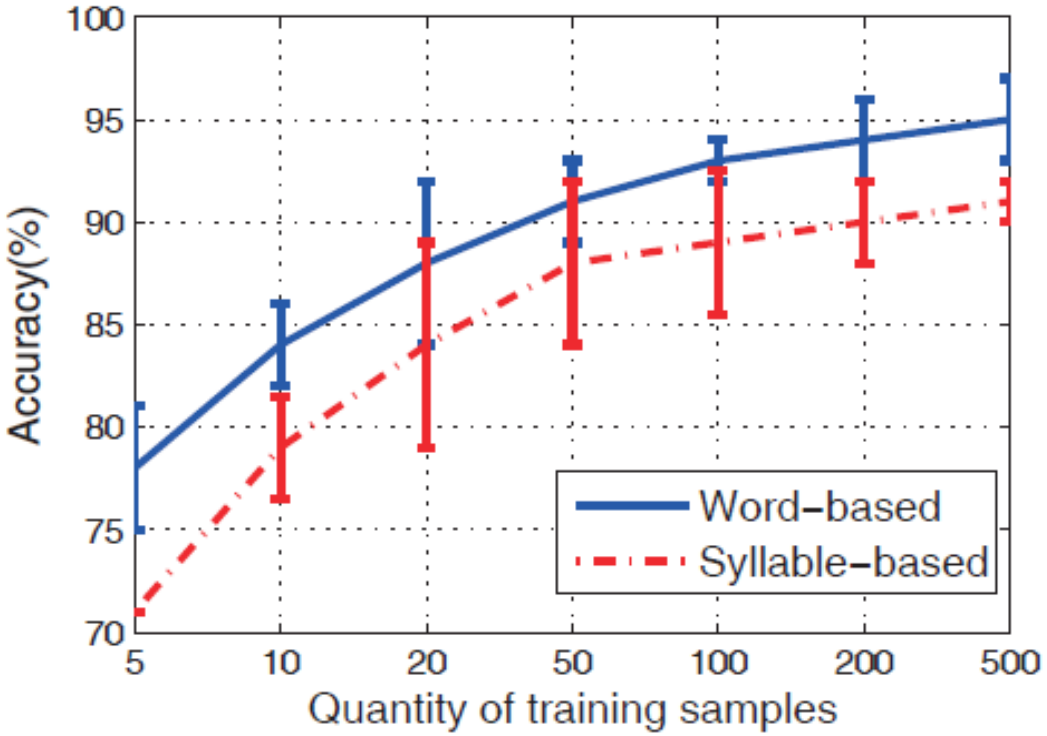
(c) Inner-word segmentation on USRP

(d) Inter-word segmentation on USRP

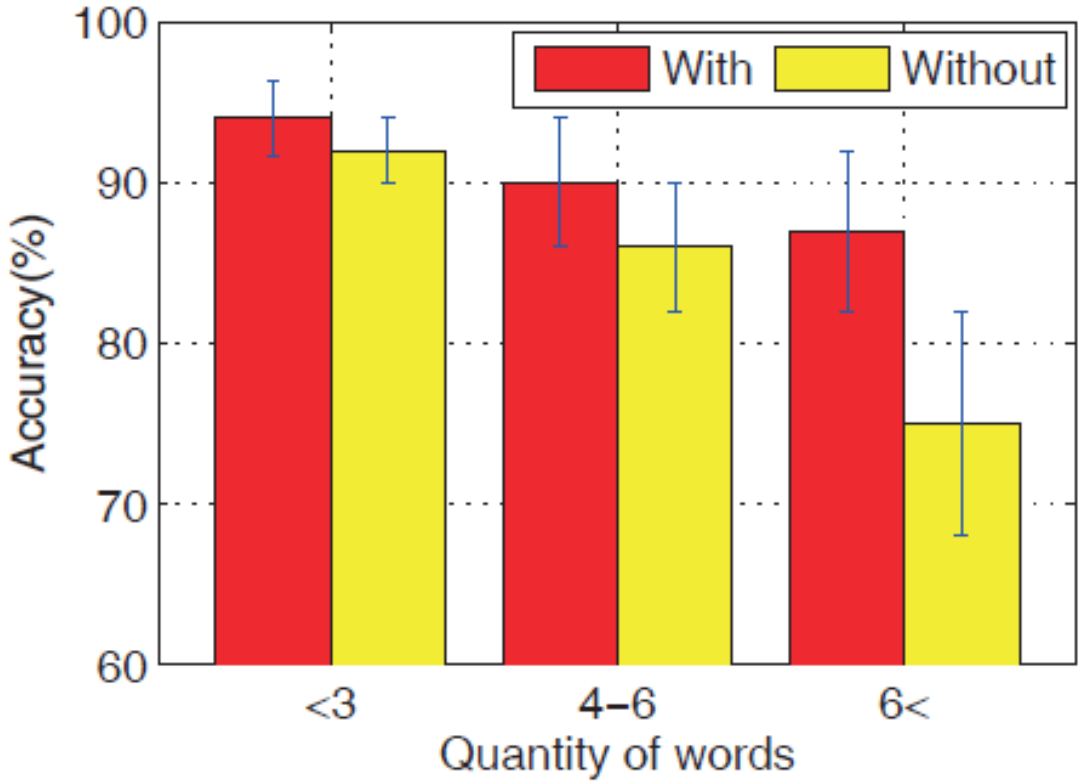
Classification Accuracy



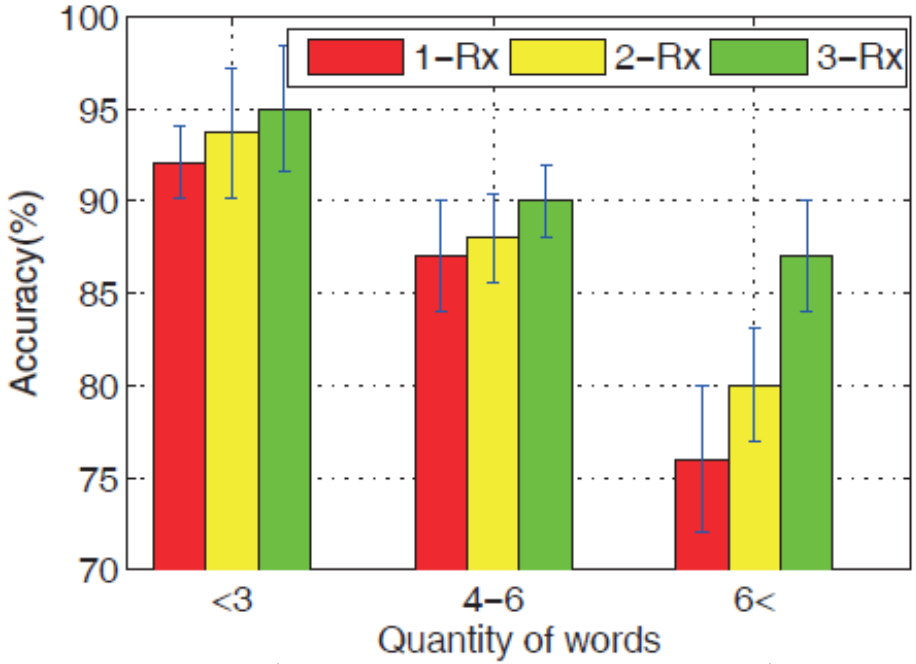
Training Overhead

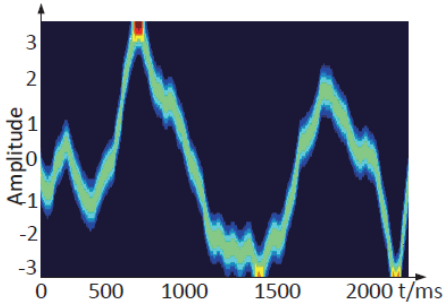


Impact of Context-based Error Correction

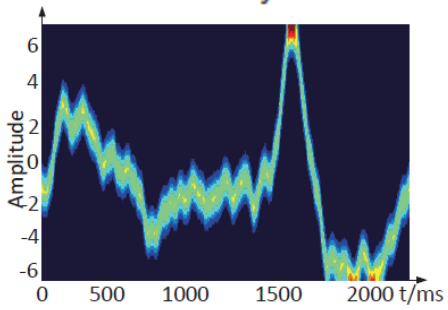


Performance with Multiple Receivers

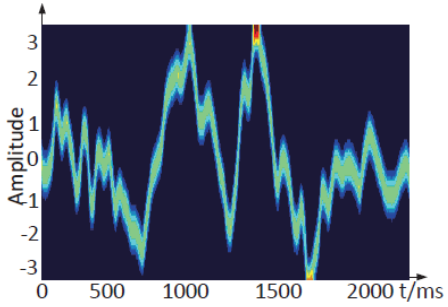




(a) GOOD 0°



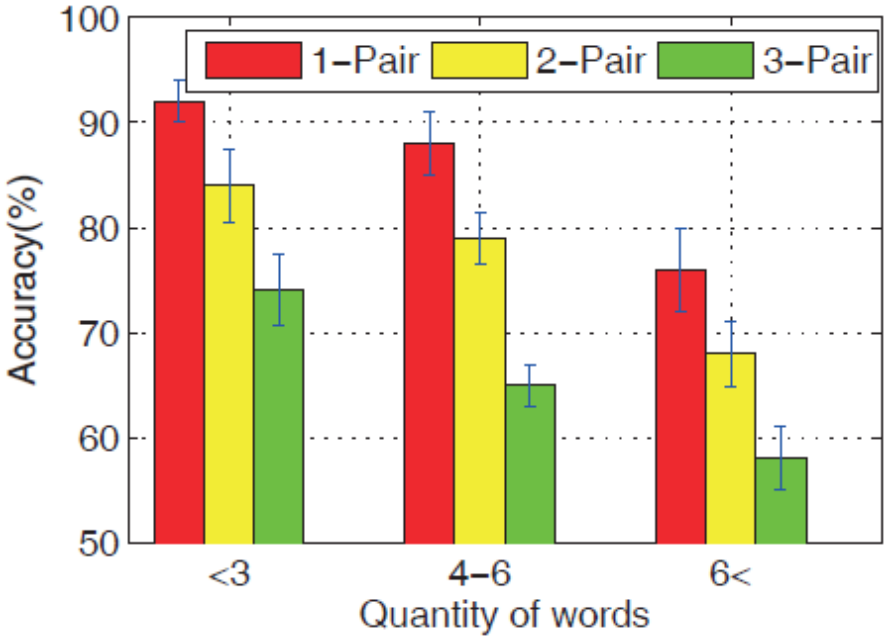
(b) GOOD 90°



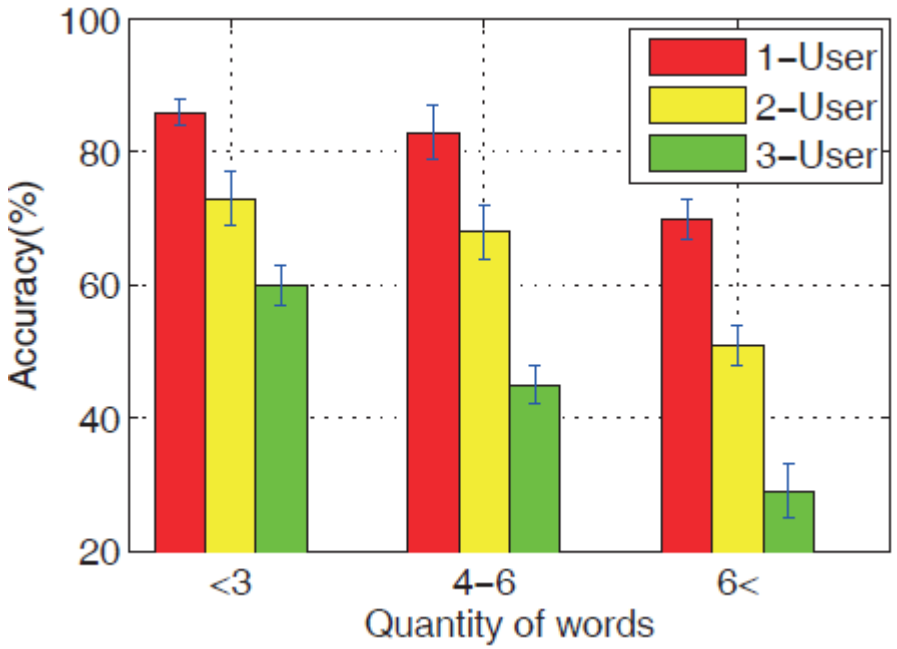
(c) GOOD 180°

Example of different views for pronouncing words

Performance for Multiple Targets

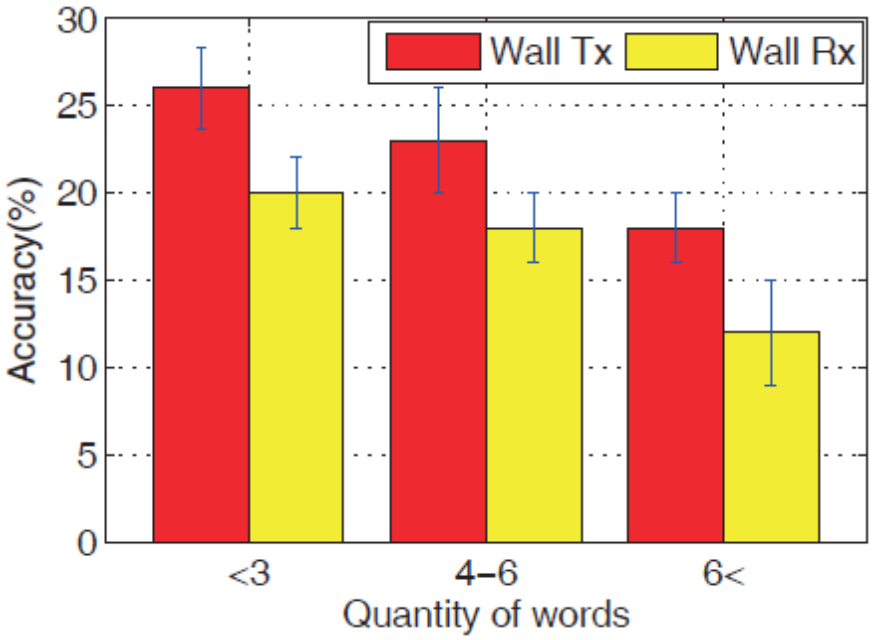


Performance of multiple users with multiple link pairs.

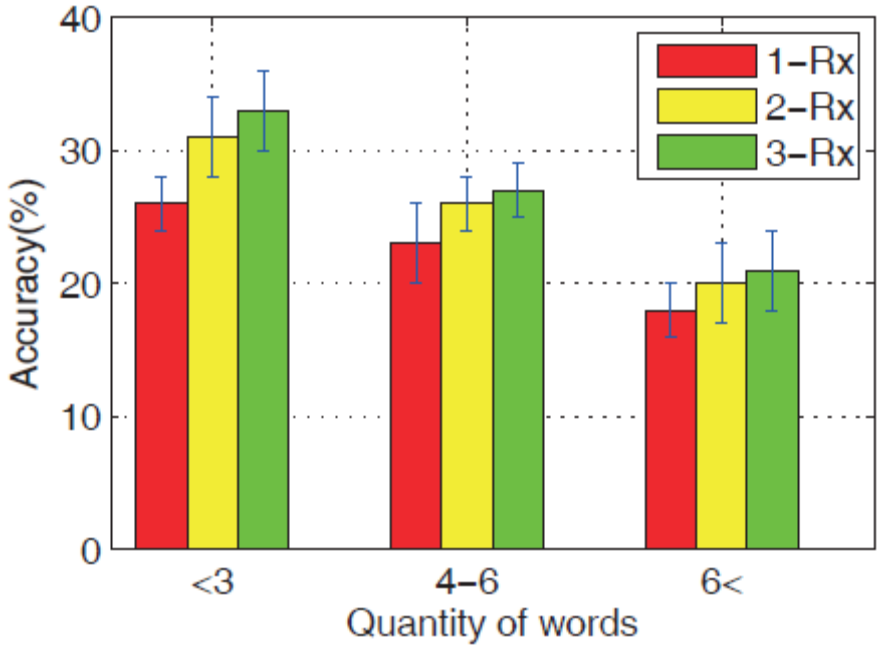


Performance of zigzag decoding for multiple users.

Through Wall Performance

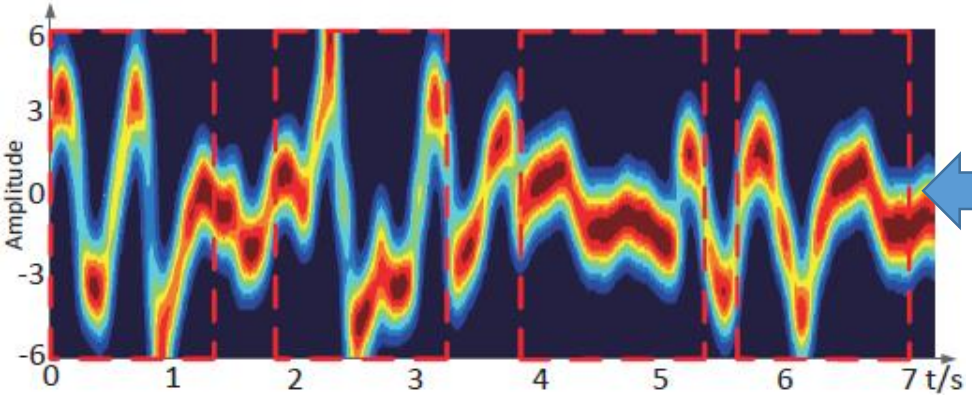


Performance of two through wall scenarios.



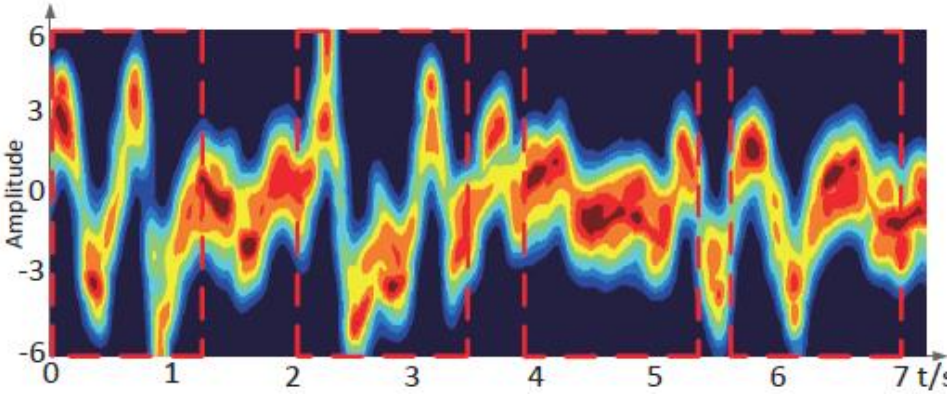
Performance of through wall with multiple Rx.

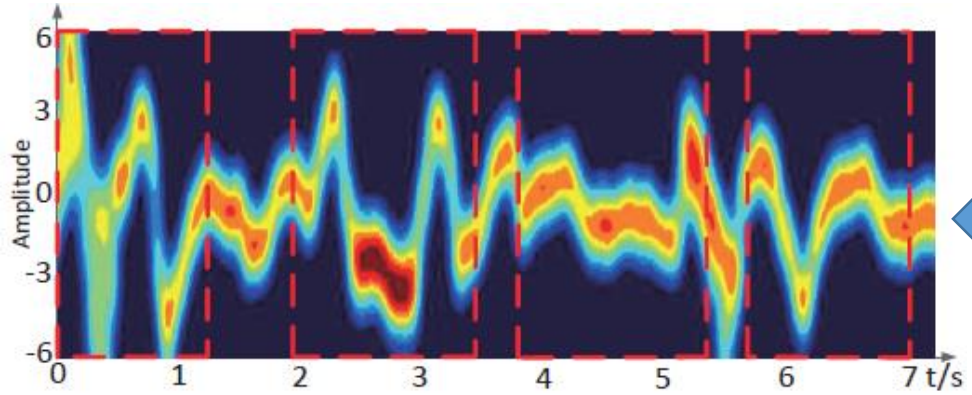
Resistance to Environmental Dynamics



Waveform of a 4-word sentence without interference of ISM band signals or irrelevant human motions

Impact of irrelevant human movements interference





Impact of ISM band interference

Conclusion

- WiHear is the 1st prototype in the world, trying to use Wi-Fi signal to sense and recognize human talks.

- WiHear takes the 1st step to bridge communication between *human speaking* and *wireless signals*.

- WiHear introduces a new way so that machine can sense more complicated *human behaviors* (e.g. mood).

Thank you for your listening !

Questions ?

WiHear

We Can Hear You With Wi-Fi !

Guanhua Wang

gwangab@cse.ust.hk