UCLA Samueli
School of Engineering

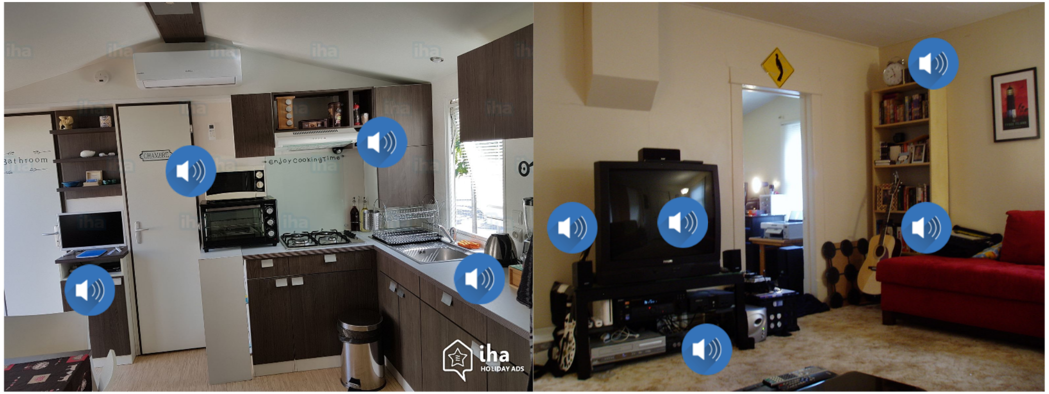
UWHear: Through-wall Extraction and Separation of Audio Vibrations Using Wireless Signals

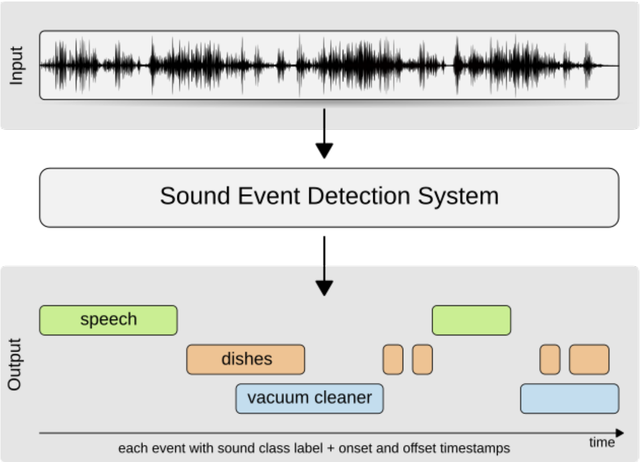
Ziqi Wang[1] , Zhe Chen[2], Akash Deep Singh[1], Luis Garcia[1], Jun Luo[2], and Mani B. Srivastava[1]
[1] University of California, Los Angeles, Electrical and Computer Engineering Department
[2] Nanyang Technological University, Singapore, School of Computer Engineering

UCLA Samueli School of Engineering

Audio Sensing and Sound Event Detection

- An ability to detect, classify, and localize complex acoustic events can be a powerful tool to help smart systems build context-awareness





A typical example of sound event detection and classification systems. Flow chart from http://dcase.community/challenge2019/task-sound-event-detection-in-domestic-environments

UCLA Samueli School of Engineering

Dual Challenges for Audio Sensing

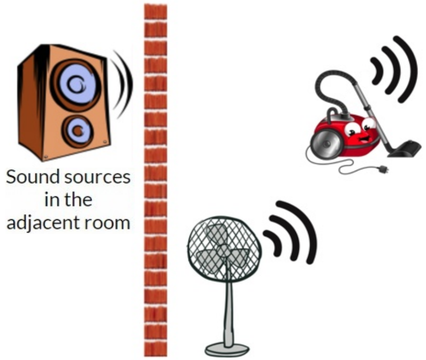
Noise / Non-target sound

Multiple Target Sounds

- Audio sensing and downstream processing are negatively affected by background noise and cross-interference between target sound sources.

Goal: Separate Sound from Multiple Sources

- A desired audio sensing system should be able to:



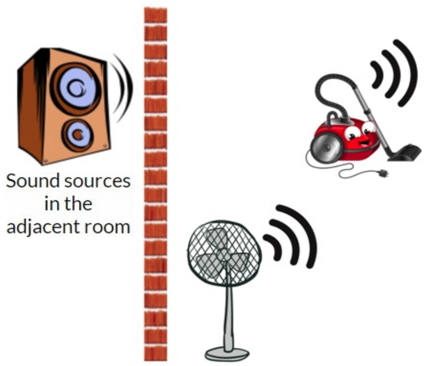
① Record audio signals from vibrations

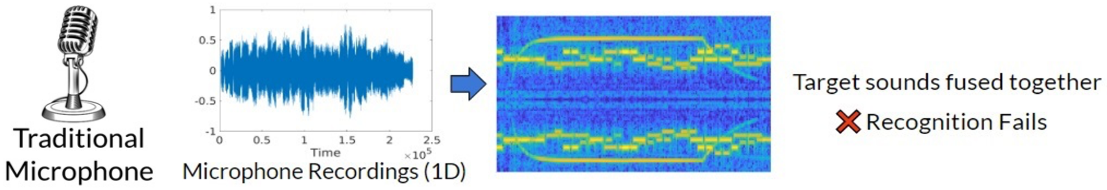
② Separate sounds from multiple sources

③ Work in non-line-of-sight (NLOS) scenarios

Audio Sensing: Microphone vs Wireless Signals

- A traditional microphone passively captures the sound pressure wave.
- Multiple target sounds and noise are blended?

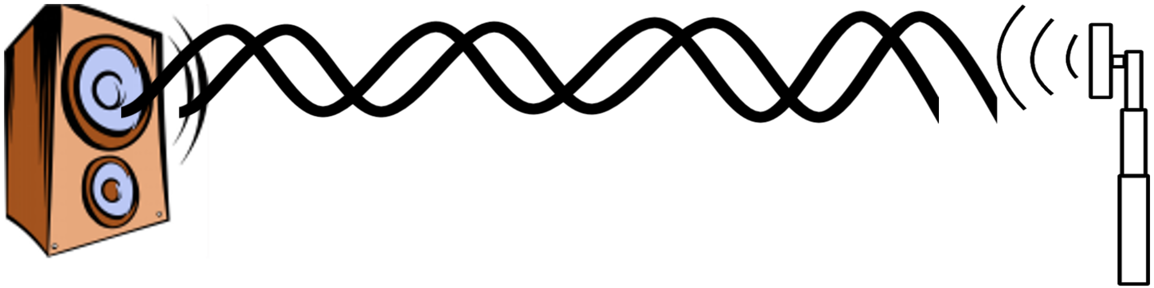




- Can we actively measure the sounds from their source vibrations?

Wireless Vibrometry

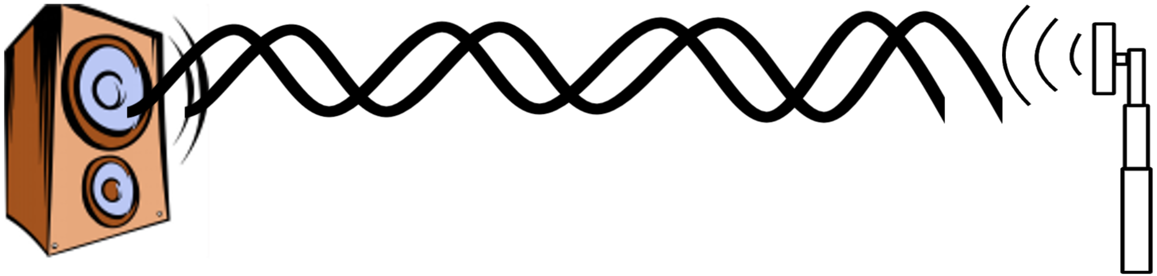
- Wireless Vibrometry is about recovering information from vibrating objects, e.g. speakers, engines…....



- Can be used to recover sound directly from its source vibrations
  - Previous works isolates one sound of interest by focusing a **highly directional beam** on its source

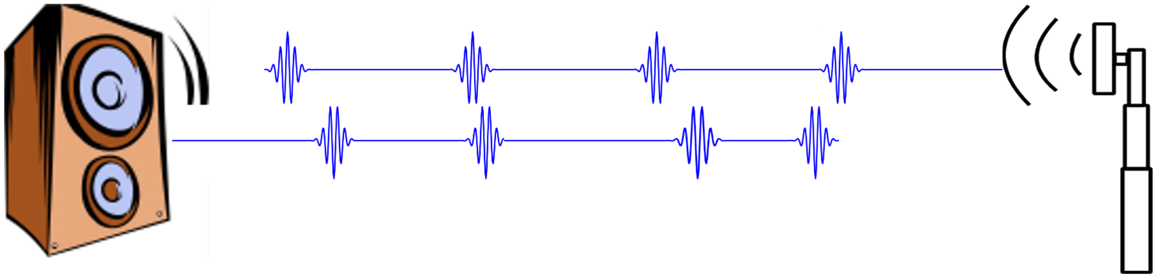
Wireless Vibrometry Using IR-UWB Radar

- Impulse-Radio Ultra Wideband (IR-UWB) works by sending out a train of very short Gaussian pulses and collect reflection pulses.



Wireless Vibrometry Using IR-UWB Radar

- Impulse-Radio Ultra Wideband (IR-UWB) works by sending out a train of very short Gaussian pulses and collect reflection pulses.



Wireless Vibrometry Using IR-UWB Radar

- Impulse-Radio Ultra Wideband (IR-UWB) Radar

Sound Separation
  - Operates with very short pulses (large bandwidth)
  - Has fine spatial resolution to separate multiple sources
  - Measures the distance to the targets with Time-of Flight

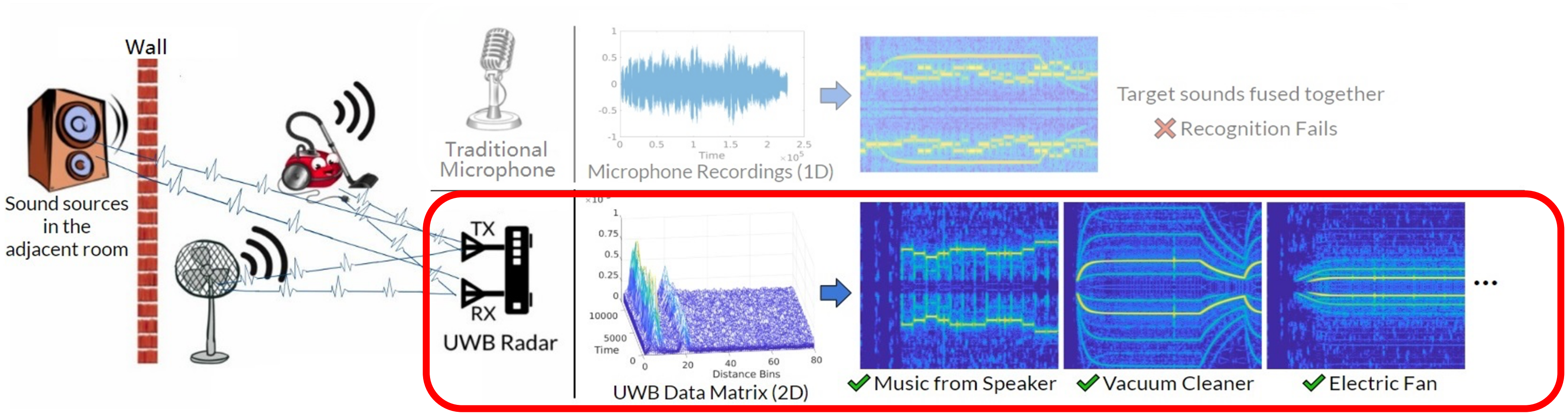
Sound Recovery
  - Can detect subtle target movements with RF phase
  - (To be shown in the next section)

NLOS
  - Works at sub-10GHz band
  - Can Penetrate light building materials

Practical
  - Is incorporated on mobile platforms (e.g., iPhone 11)
  - Is low-cost and has low power consumption

Audio Sensing: Microphone vs Wireless Signals

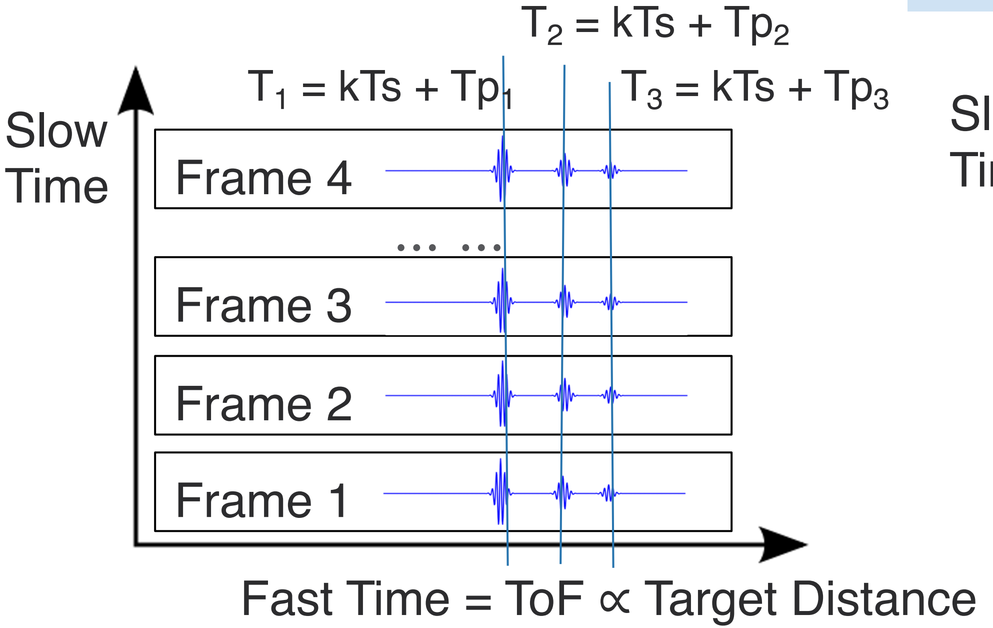
- **UWHear** is a system that uses Impulse Radio Ultra-Wideband (IR-UWB) to separate and recover sounds from multiple sources simultaneously.



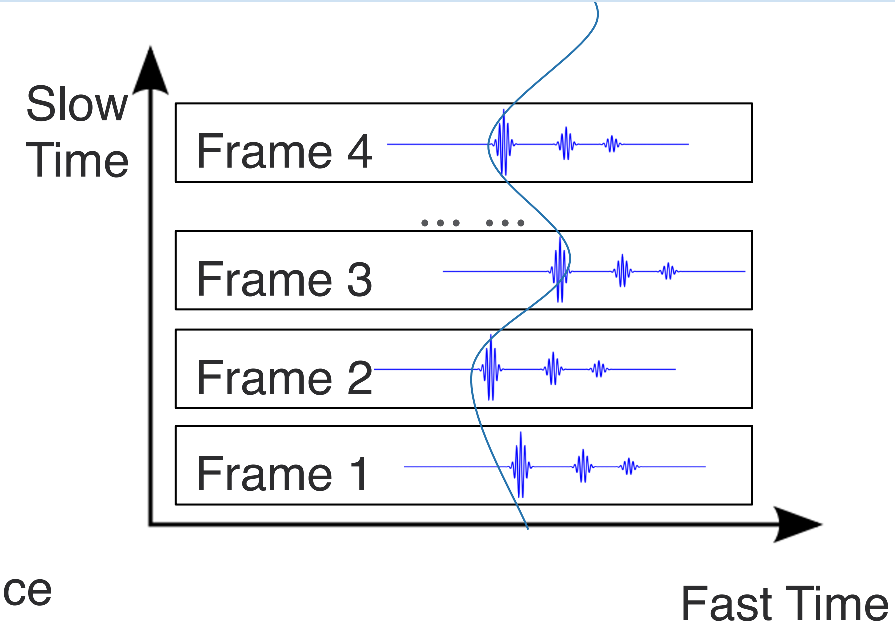
UCLA Samueli School of Engineering

IR-UWB Audio Sensing Theory

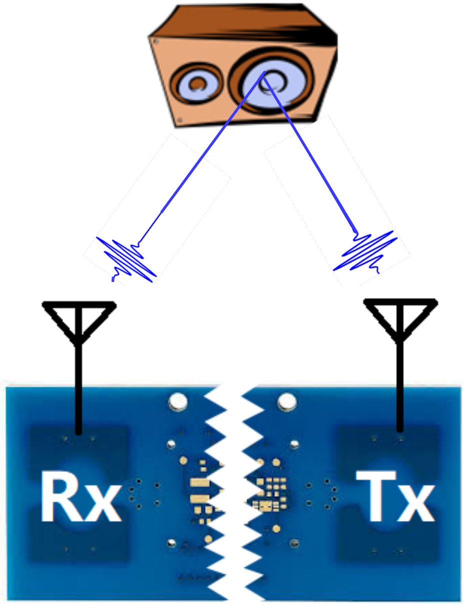
Data Structure: Fast time and Slow time

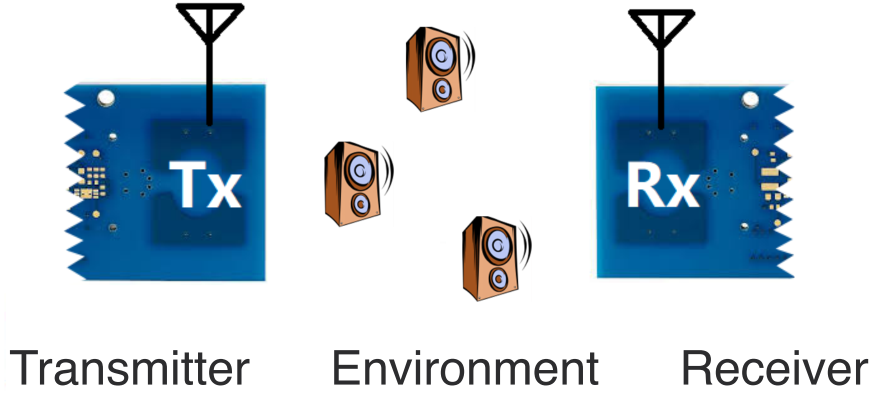


Challenge: Speaker displacement ~mm level, UWB spatial resolution ~cm level

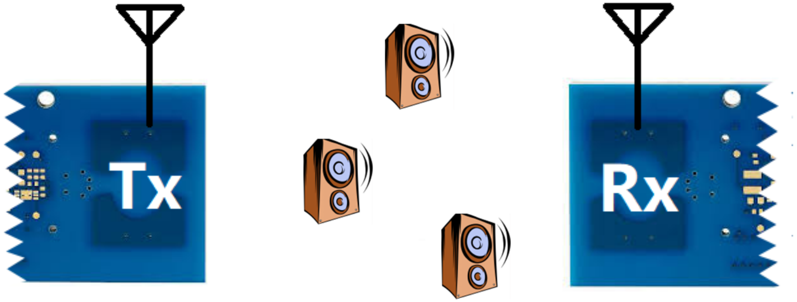


IR-UWB Audio Sensing Theory

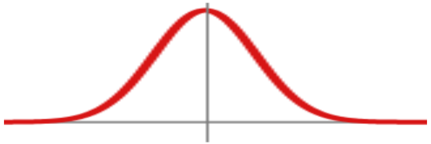




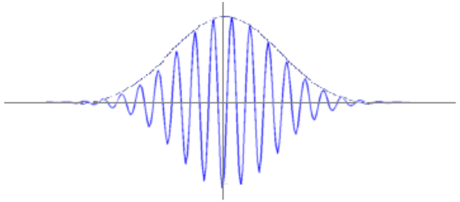
IR-UWB Audio Sensing Theory



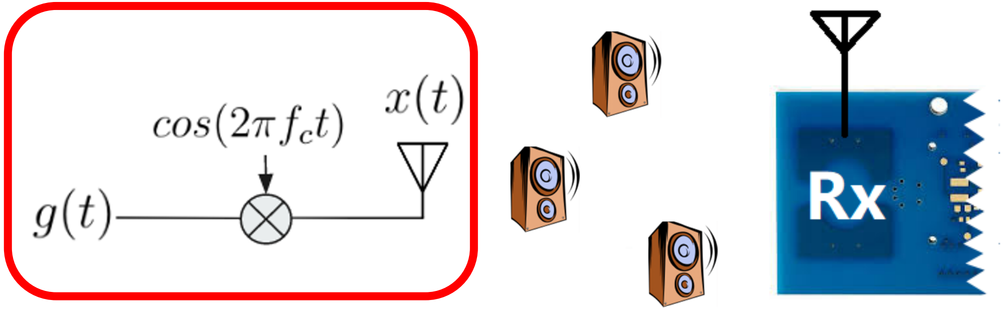
Baseband Gaussian pulse: $g(t)$



Transmitter Output: $x(t) = g(t - kT_s)cos(2\pi f_c(t - kT_s)),$

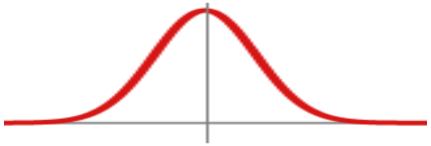


IR-UWB Audio Sensing Theory

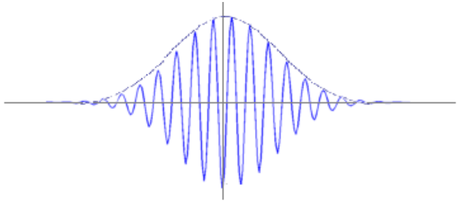


Transmitter

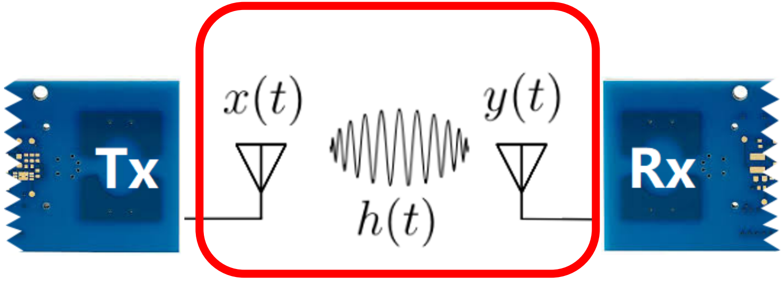
Baseband Gaussian pulse: $g(t)$



Transmitter Output: $x(t) = g(t - kT_s)cos(2\pi f_c(t - kT_s)),$



IR-UWB Audio Sensing Theory



Environment

Environment Response

Caused by the placement distance of speaker p

$$h(t) = \sum_{p=1}^{P} \alpha_p \delta(t - \boxed{T_p} - \boxed{T_p^D(t)})$$

Caused by vibration of speaker p

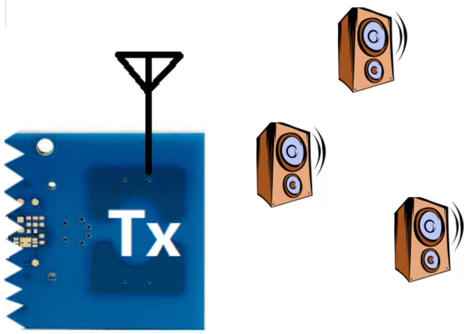
RF Received Signal

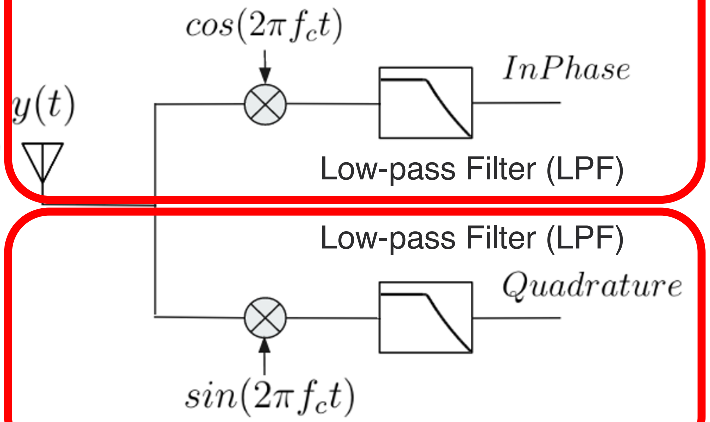
$$y(t) = x(t) * h(t) + n(t)$$

The **phase** of received RF signal is proportional to **speaker vibration**

$$= \sum_{p=1}^{P} \alpha_p g(t - kT_s - T_p - T_p^D(t)) cos(2\pi f_c(t - kT_s - T_p - \boxed{T_p^D(t)})) + n(t).$$

IR-UWB Audio Sensing Theory





Down Conversion (Demodulate) to Baseband:

$$y_{in-phase}(t) = LPF[y(t) \times cos(2\pi f_c(t - kT_s))]$$

$$y_{quad}(t) = LPF[y(t) \times sin(2\pi f_c(t - kT_s))]$$

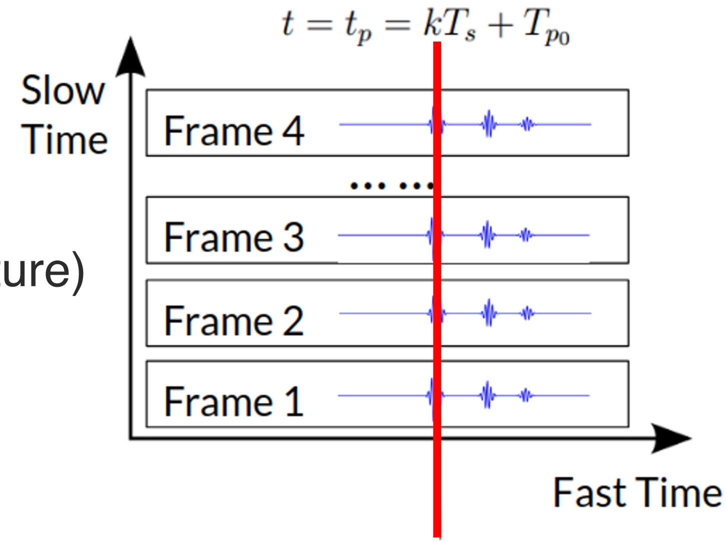
Apply Sound Source Separation using ToF

Apply sound source separation

Select a particular sound source by fixing $t = t_p = kT_s + \boxed{T_{p_0}}$

ToF caused by the placement distance of speaker p_0

Data Matrix
(In-phase or Quadrature)



Linear Approximation

- The I/Q data $\approx V_{\text{tx}}$

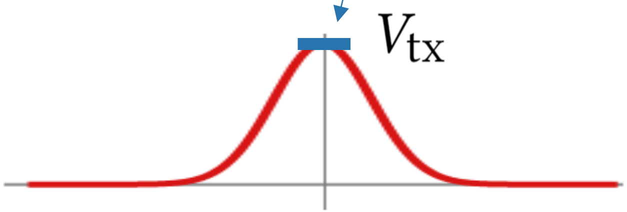
Time-varying phase offset caused by speaker p_0 vibration

$$y_{in-phase}(t_p) = \frac{1}{2}\alpha_{p0}\boxed{g(T_{p0}^D(t_p))}cos(\boxed{2\pi f_c T_{p0}} + \boxed{2\pi f_c T_{p0}^D(t_p)}) + \tilde{n}(t_p).$$

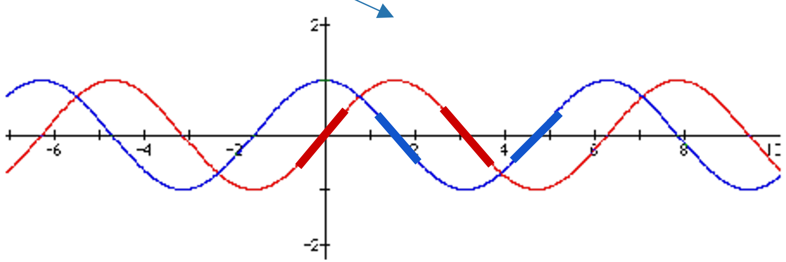
Constant phase offset caused by the distance of speaker p_0

$$y_{quad}(t_p) = \frac{1}{2}\alpha_{p0} g(T_{p0}^D(t_p))sin(2\pi f_c T_{p0} + 2\pi f_c T_{p0}^D(t_p)) + \tilde{n}(t_p).$$

- Linear approximation



Gaussian Pulse

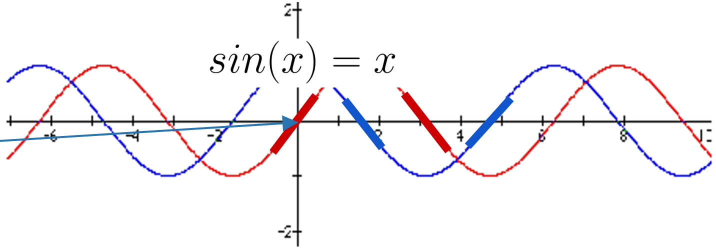


Sine and Cosine Function

IR-UWB Audio Sensing Theory

$$y_{\text{in-phase}}(t_p) = \frac{1}{2}\alpha_{p_0} V_{\text{tx}} \cos\left(2\pi f_c T_{p_0} + 2\pi f_c T^D_{p_0}(t_p)\right)$$

$$y_{\text{quad}}(t_p) = \frac{1}{2}\alpha_{p_0} V_{\text{tx}} \sin\left(2\pi f_c T_{p_0} + \underline{2\pi f_c T^D_{p_0}(t_p)}\right)$$



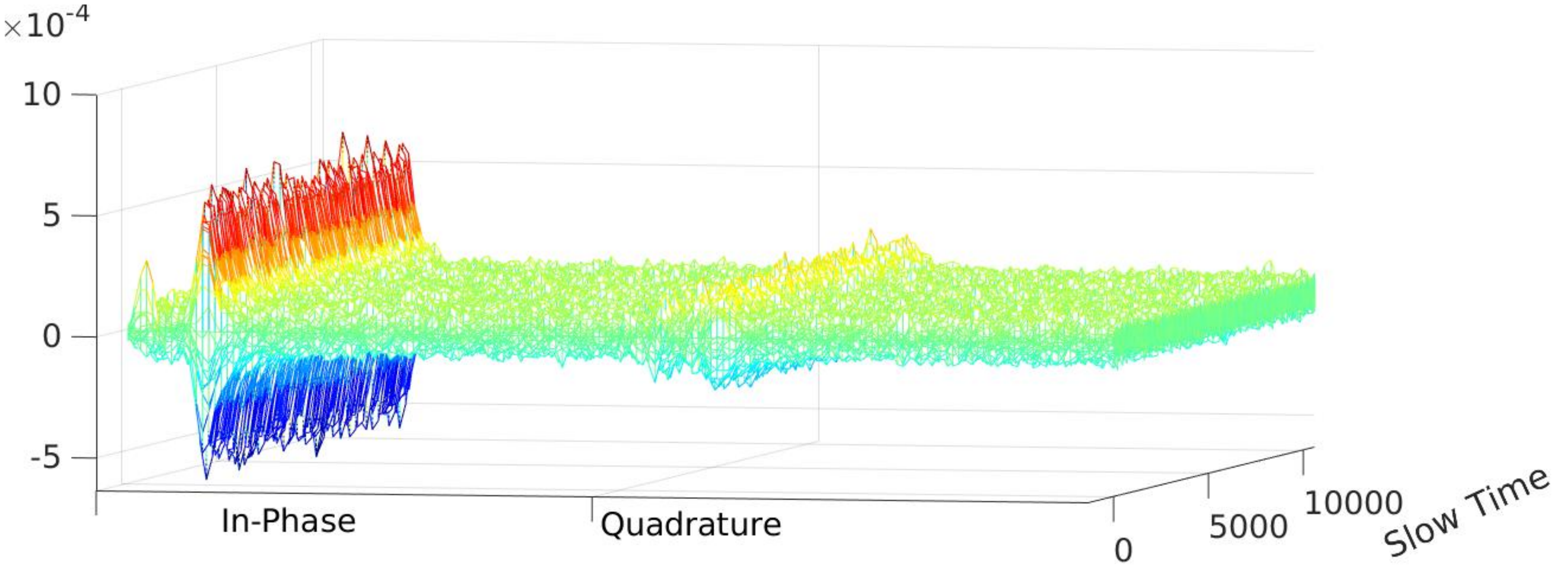
- Linear approximation

- Example: $mod(2\pi f_c T_{p_0}, 2\pi) \approx 0$

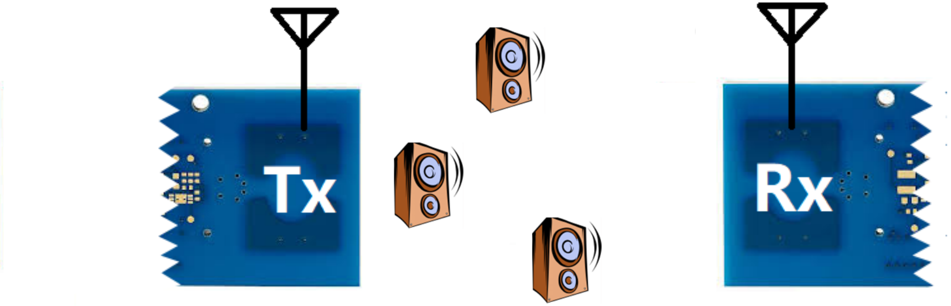
$$y_{quad}(t_p) = \frac{1}{2}\alpha_{p_0} V_{TX} 2\pi f_c T^D_{p_0}(t_p)$$

- In this example, the **speaker vibration movement** is proportional to **quadrature amplitude** change
- In other cases, the speaker displacement is proportional to **in-phase** amplitude change

IR-UWB Audio Sensing Theory



IR-UWB Audio Sensing Theory



Transmitter Environment Receiver

Gaussian Pulse → Transmitted RF Pulse → Received RF Pulse → Baseband I/Q Signal

Subtle ToF Change

Time-varying Phase

In-phase/Quadrature Amplitude

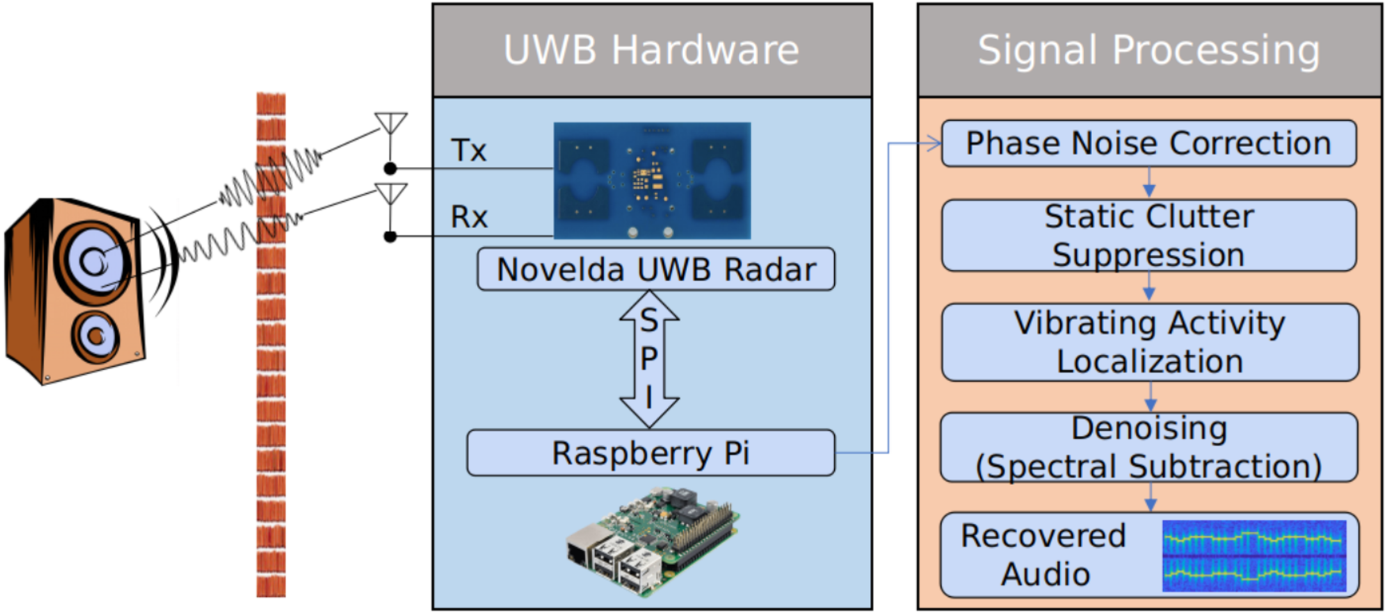
IR-UWB Signal

Vibration-related Information

UCLA Samueli School of Engineering

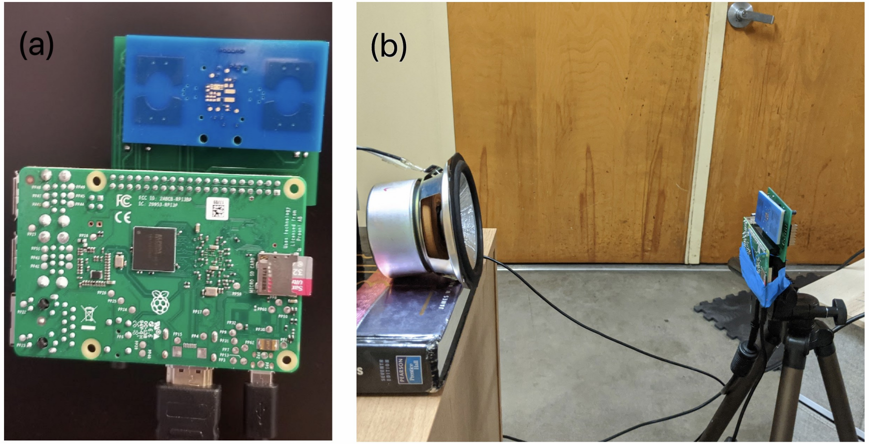
UWHear System Design

System Overview



Hardware Implementation

- XeThru X4M05 UWB Radar Sensor
- Controlled and sending data to Raspberry Pi 3B+ via SPI interface
- Gaussian pulses with 1.4 GHz bandwidth centered at 7.29GHz
- Can comply with FCC regulations, Sampling rate 1.5 kHz



Signal Processing Pipeline

Raw UWB Data

Phase Noise Correction

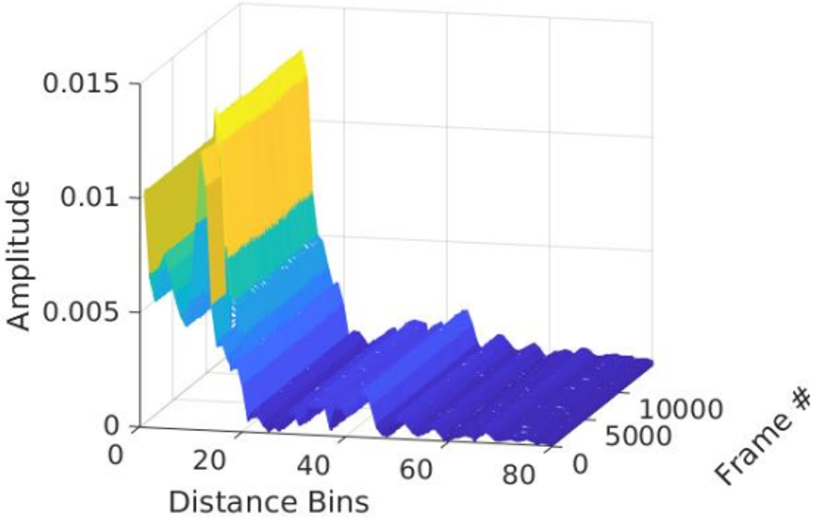
Static Clutter Suppression

Vibrating Target Localization

Spectral Subtraction

Recovered Sound

Further Analysis



Signal Processing Pipeline

Raw UWB Data

Phase Noise Correction

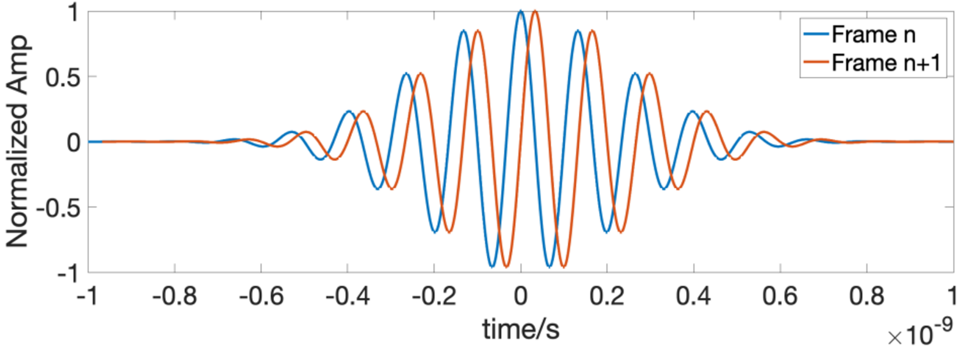
Static Clutter Suppression

Vibrating Target Localization

Spectral Subtraction

Recovered Sound

Further Analysis



Signal Processing Pipeline

Raw UWB Data

Phase Noise Correction

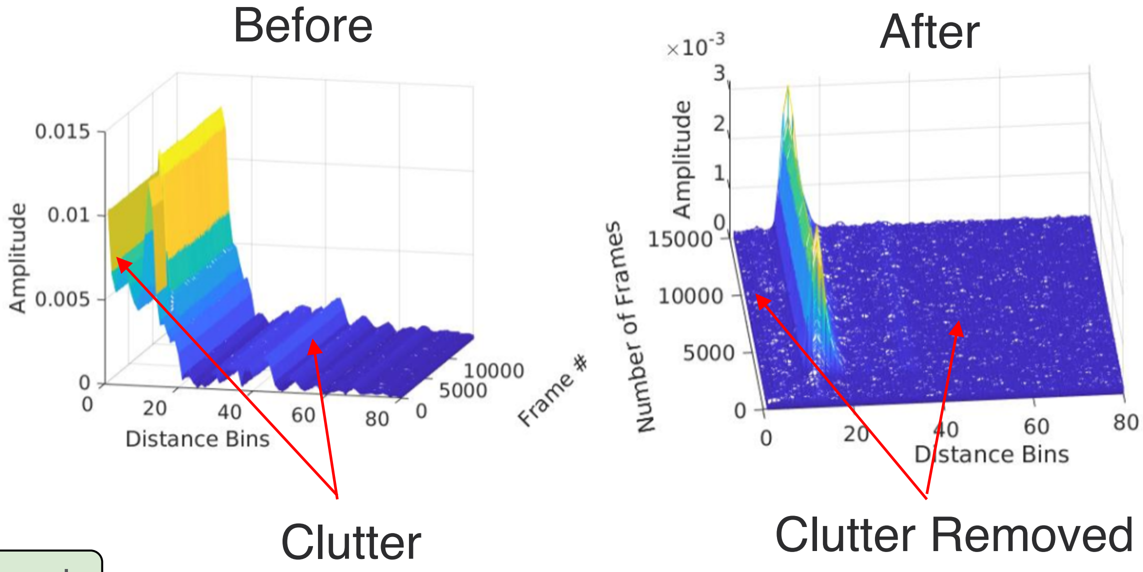
Static Clutter Suppression

Vibrating Target Localization

Spectral Subtraction

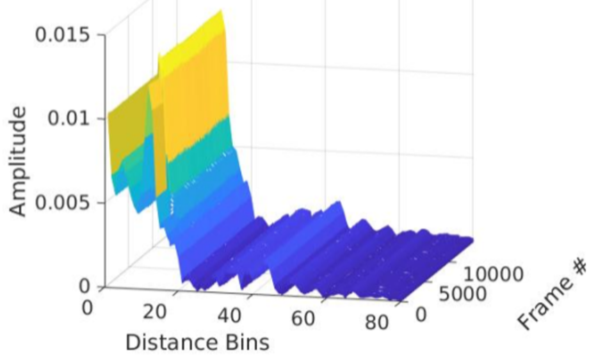
Recovered Sound

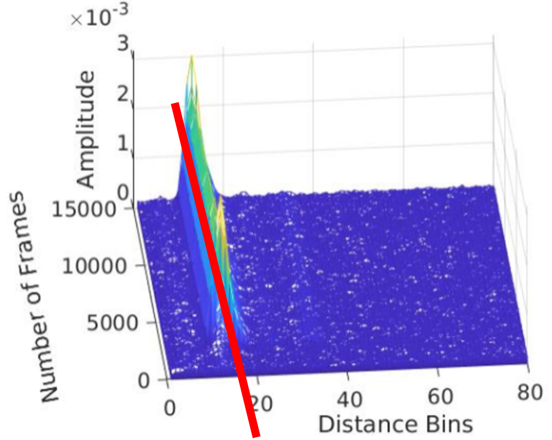
Further Analysis



Signal Processing Pipeline

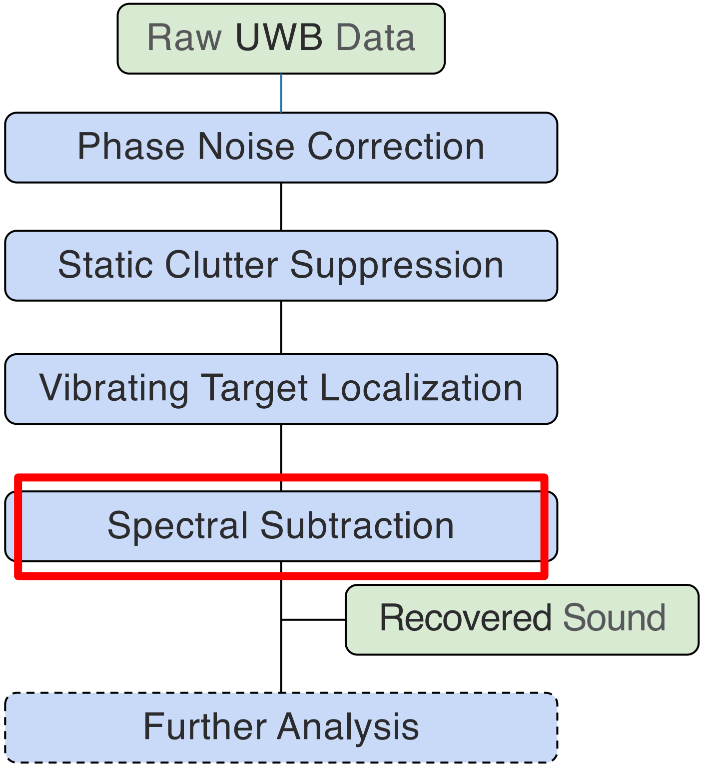
- Raw UWB Data
- Phase Noise Correction
- Static Clutter Suppression
- **Vibrating Target Localization**
- Spectral Subtraction
  - Recovered Sound
- Further Analysis

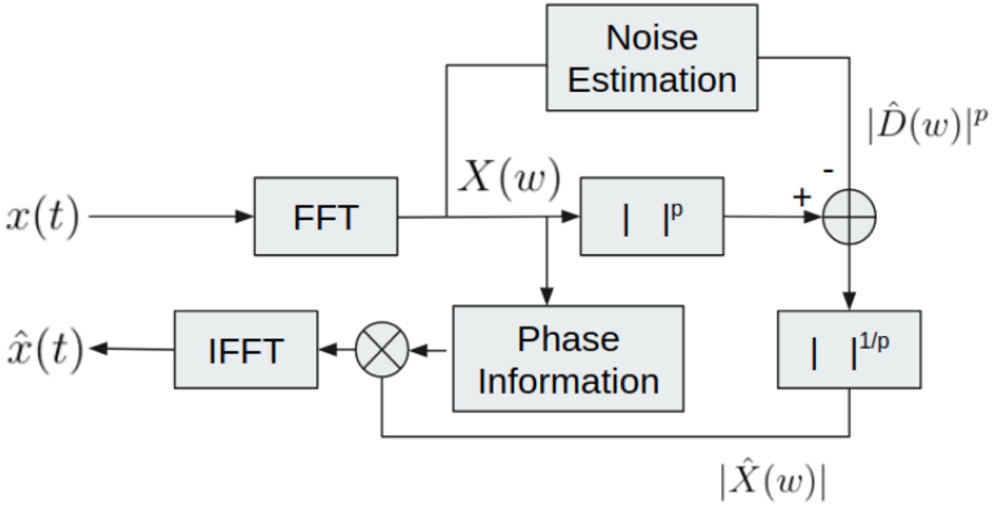




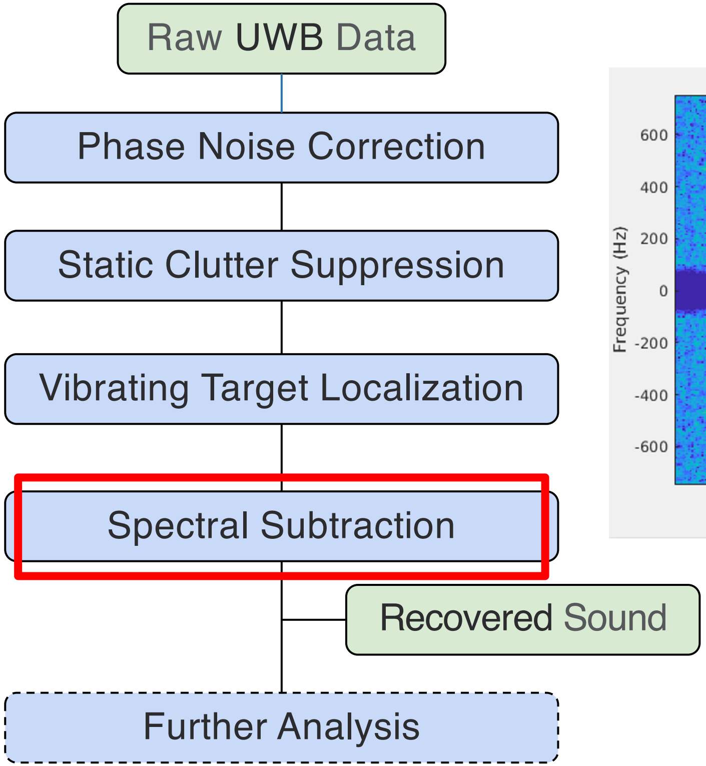
Distance = 1m

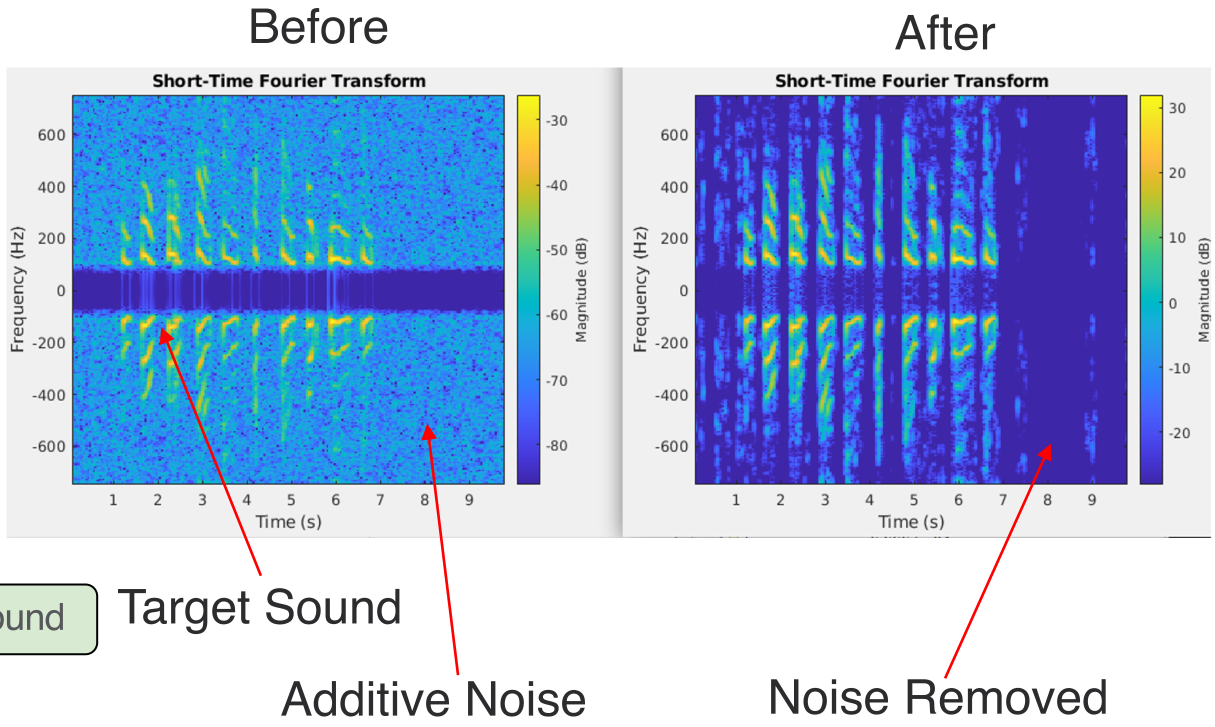
Signal Processing Pipeline





Signal Processing Pipeline





Signal Processing Pipeline

Raw UWB Data

Phase Noise Correction

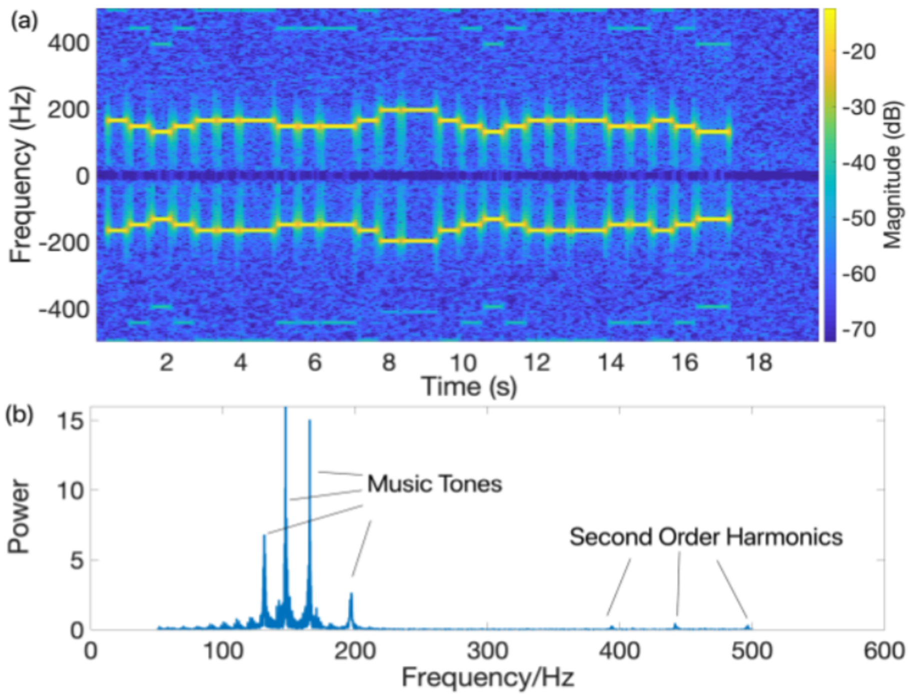
Static Clutter Suppression

Vibrating Target Localization

Spectral Subtraction

Recovered Sound

Further Analysis

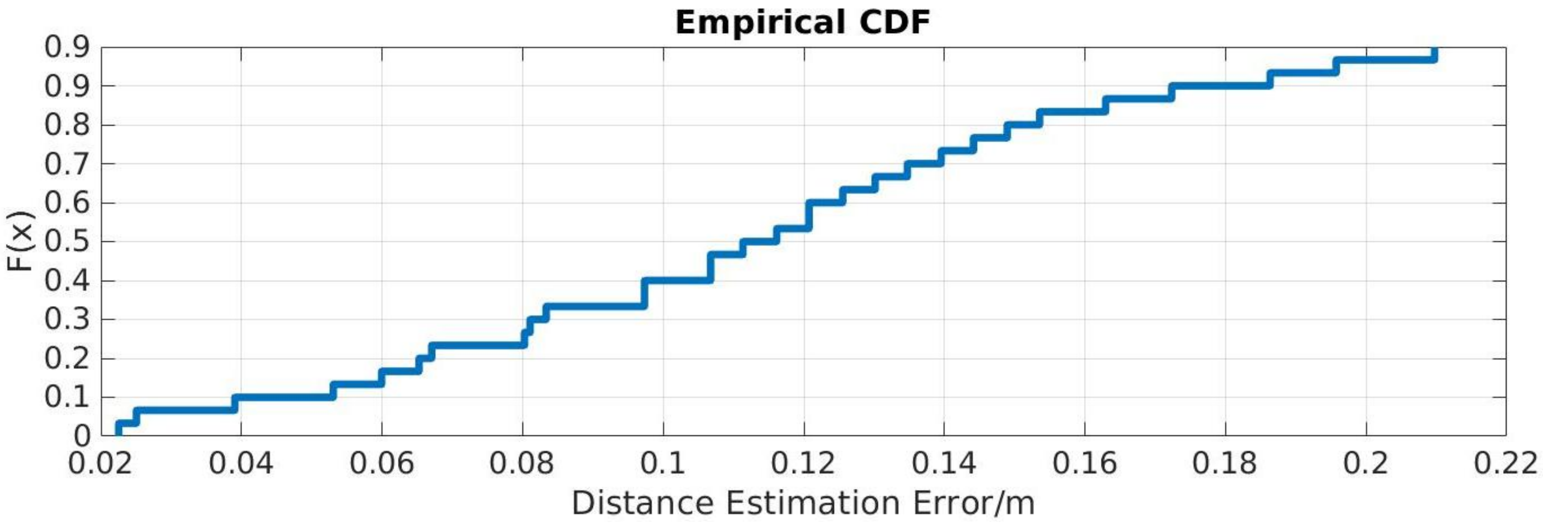


UCLA Samueli School of Engineering

Results

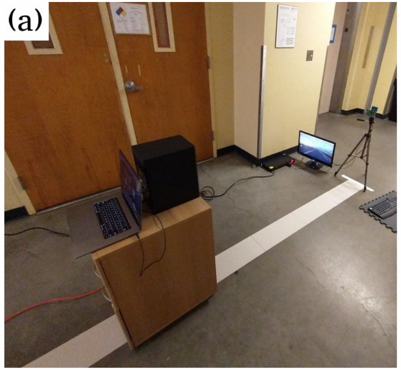
Distance to Sound Sources

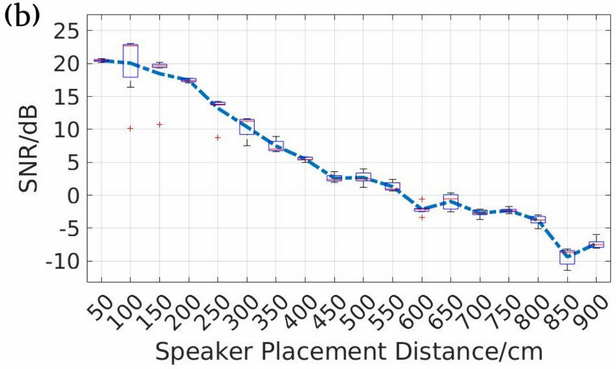
- Measured the distance to the sound sources
- The mean error is 11.19cm, the median error is 11.37cm, σ=4.88cm.

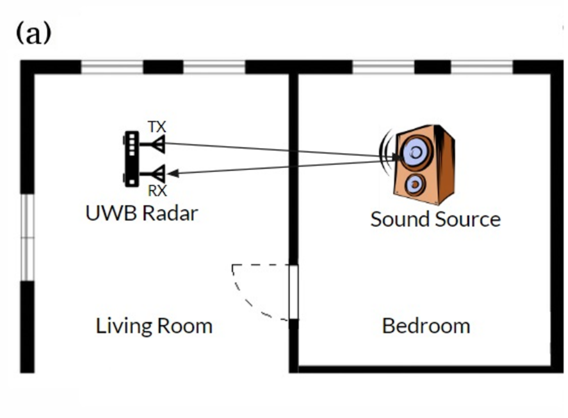


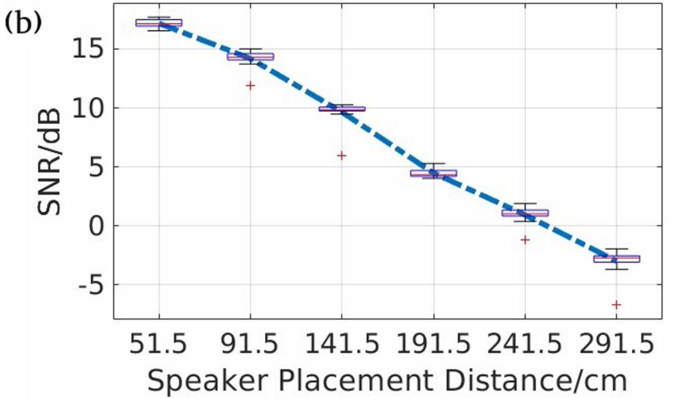
Effective Range

- 6-8m in Free Space
- 2.5m through a wooden wall



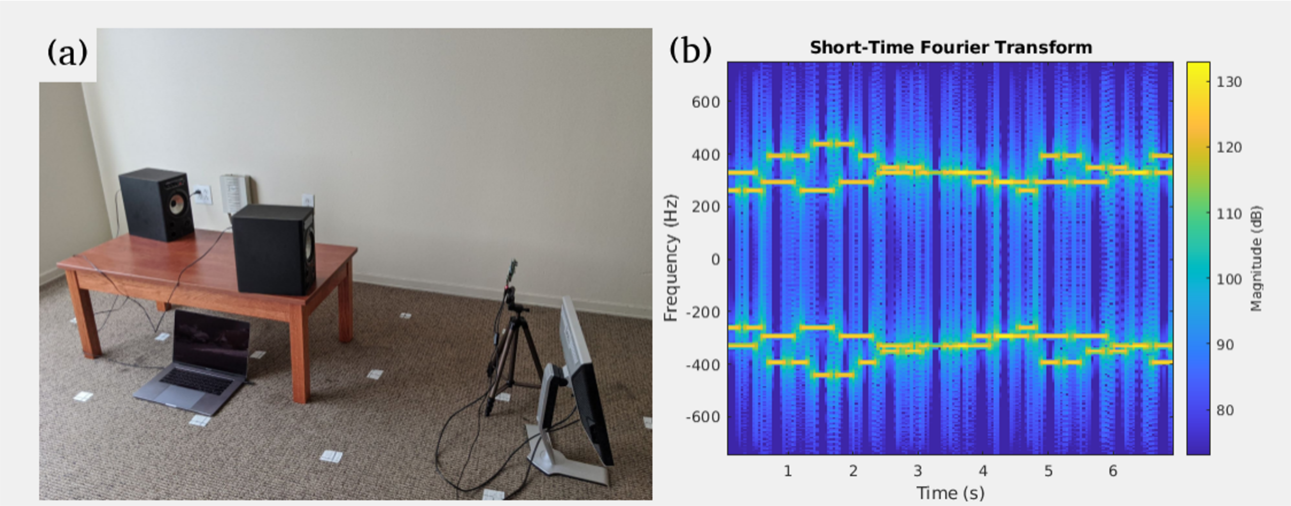




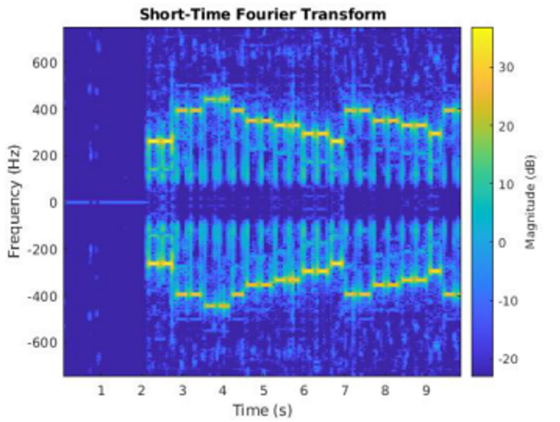


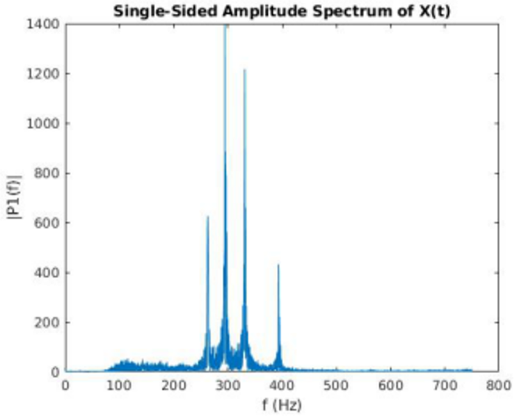
Demo: Sound Separation

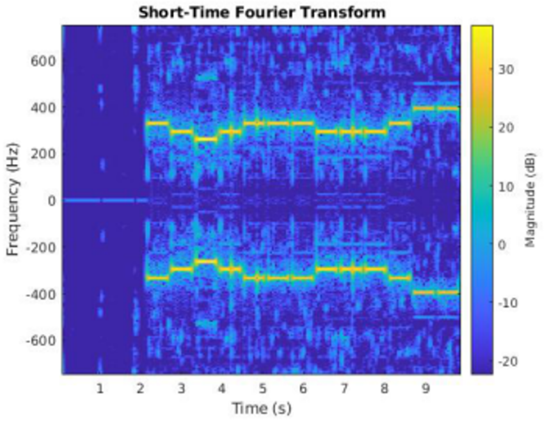
- Two speakers put at different distances and playing different contents
- One at 58cm playing *Mary has a little lamb*
- The other at 122cm *playing Twinkle twinkle little star*
- Please see the demo on the next page

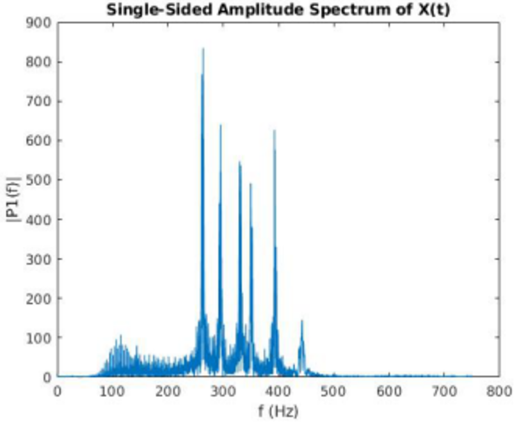


Further results demonstrate that if two sources are placed **25 cm** apart, their sounds can be recovered separately without any cross-interference.



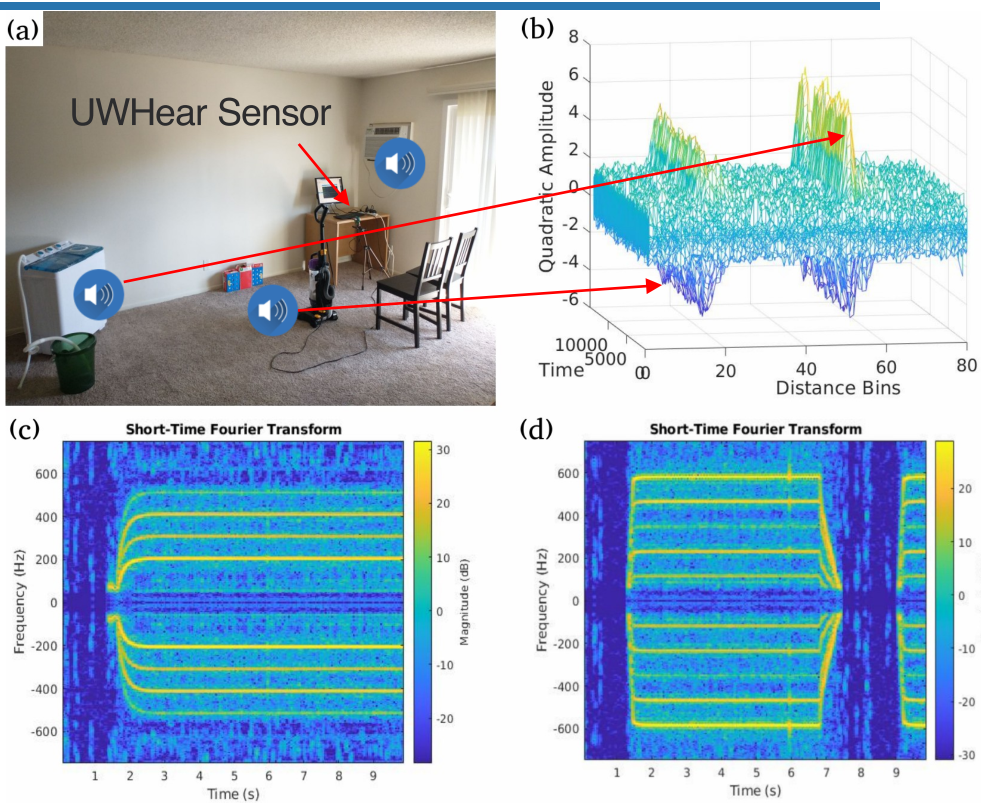






Dealing with Heterogeneous Sound Sources

- We tested UWHear in a more natural household setting
- Sound Source: Washing Machine, Vacuum Cleaner
- Noise Source: Wall AC Unit



Vacuum Cleaner

Washing Machine

Limitations and Future Work

- Increase Sampling Rate:
  - The current sampling rate (Fs = 1.5kHz) causes low sound quality
  - Fs is mainly limited by the SPI interface transmission speed
  - Using FPGA as the host may help

- Increasing Field-of-View (FoV)
  - Current directional antenna gives 50 degrees of FoV
  - Novel hardware design may use omnidirectional antenna / MIMO

- Multi-model Sensing Platform
  - UWHear cannot retrieve sound from human throat directly
  - Multiple Wireless Signals at different bands can compensate for each other
  - IR-UWB has more potentials then sensing vibration

Thank you!

SRC® JUMP CONIX ARL NSF

UCLA NESL Networked & Embedded Systems Laboratory

NANYANG TECHNOLOGICAL UNIVERSITY SINGAPORE