

# GraspTracker:

Tracking user grab posture  
with built-in sensors on mobile devices

---

SeungJoo Lee, HyungJun Yoon

2019.12.12

# Reminder: Our mid-presentation

## Problem

- **One-handed interaction** is common, but **slow** and **uncomfortable**
- Research and patents about one-handed interface **but no way to track**

**Real-time** tracking of  
user's **hand posture** with only **built-in sensors**

# Previous approach: detecting transitions



**Detect unique motion feature** with accelerometer & gyroscope

# Using inaudible sound



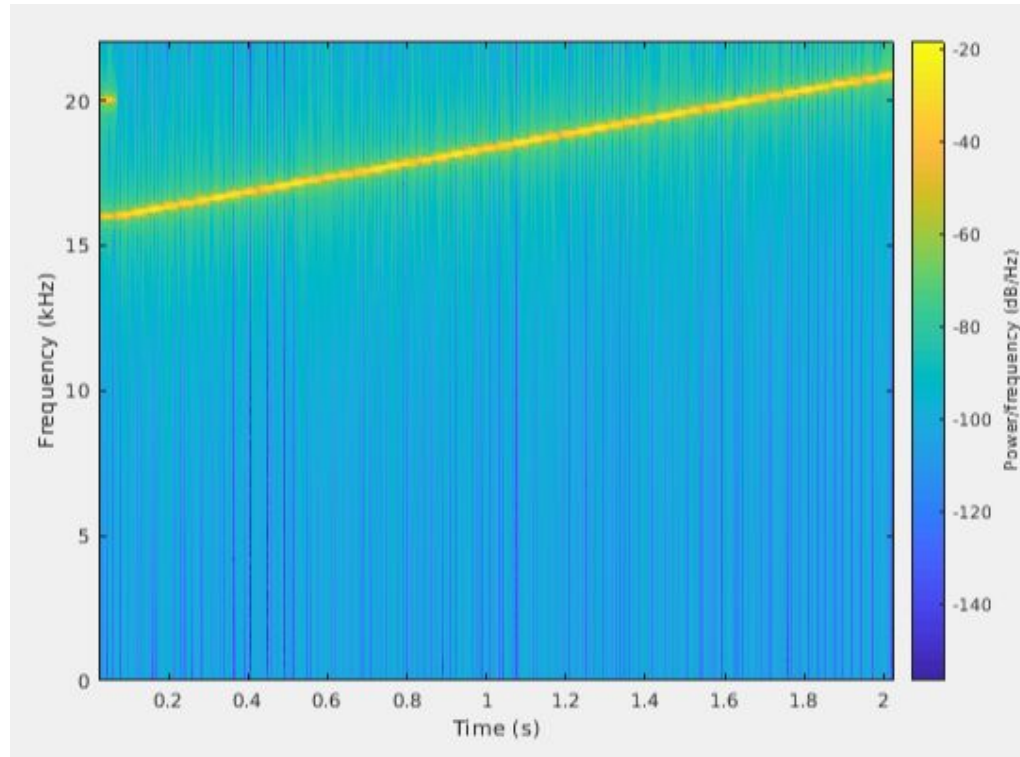
- Direct sound
  - Device body
  - Air
- Reflected sound
  - Reflected by the surrounding environments

# Direct sound

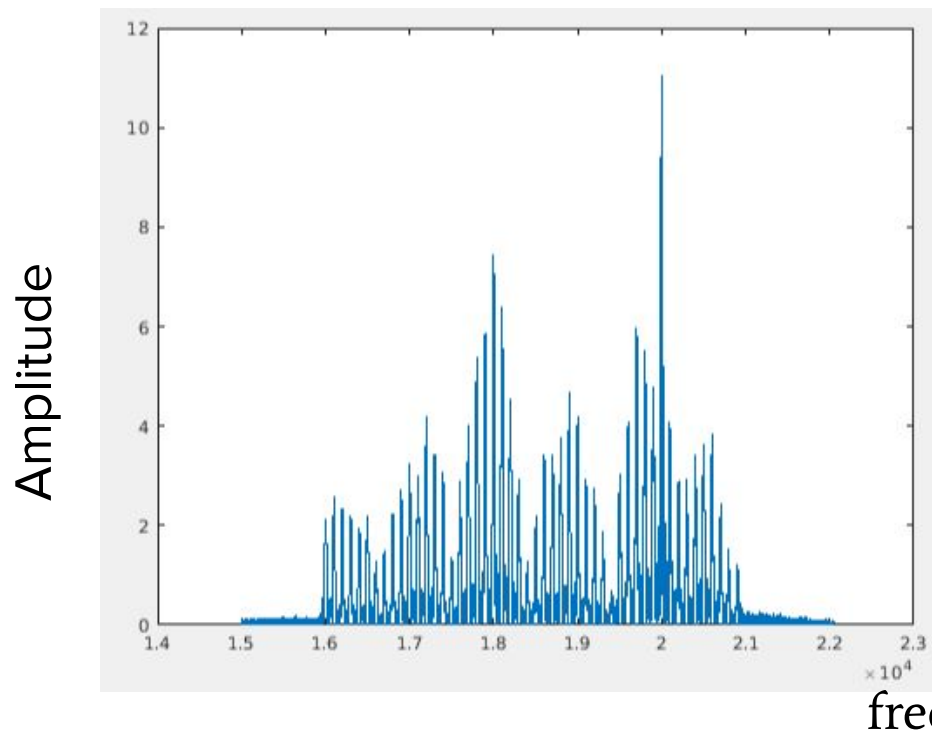
- Device body
  - Affected by the contact caused by the grip
    - Damping conditions
    - Differs for each frequency
- Air
  - Affected by the skin of the hand
    - Human hand absorbs the acoustic signal
    - Absorption rate differs for each frequency

Generate sound with **various frequencies** and **observe the response**

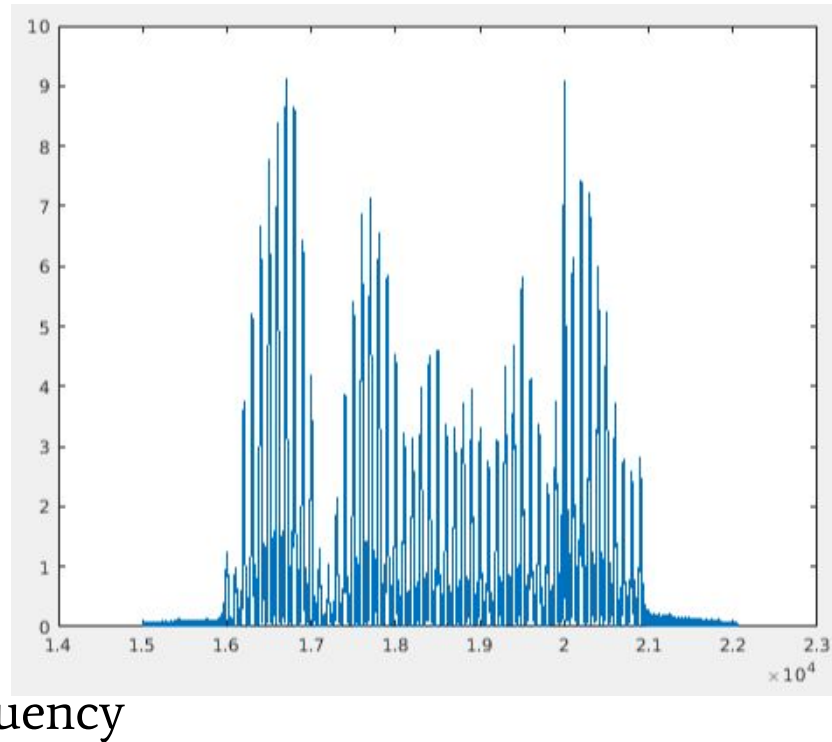
# Sound for testing



# Response from different grasp pose



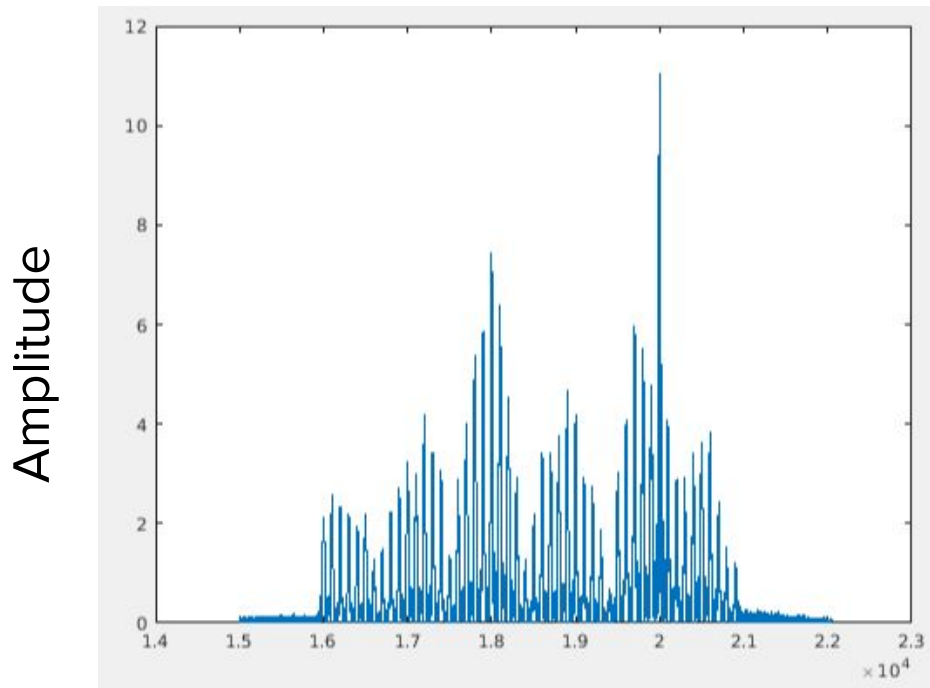
Two hand



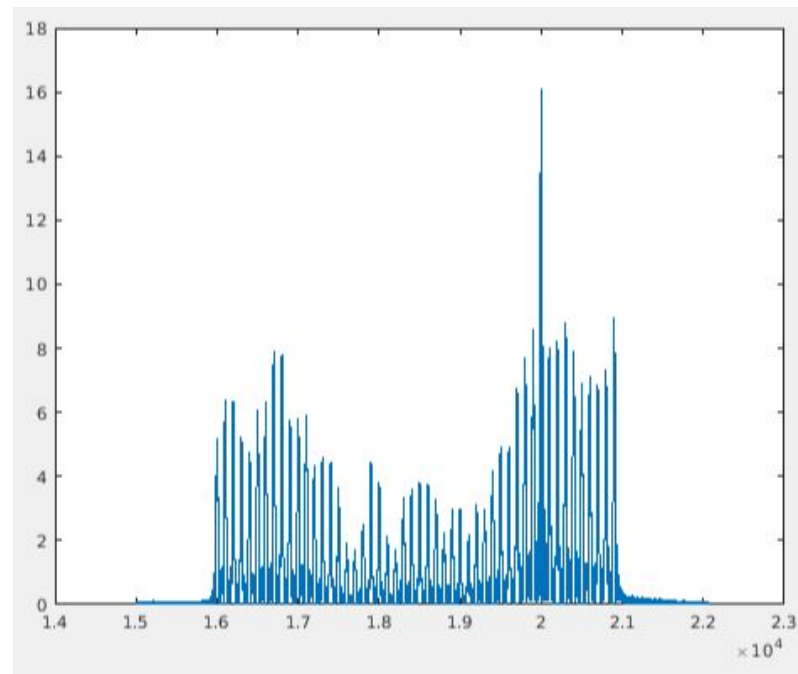
Left hand

# Reflected sound

- Information about surrounding environment



Two hand - env1



Two hand - env2



# Considering only direct sound

- Difference in speed of sound in air and solid

Medium (20 °C)	Speed of Sound Waves (m/s)
Dry Air	343
Water	1437
Wood	3850
Glass	4540
Aluminum	6320

- Smartphone with aluminum, 10cm
  - Time for the first sample comes through air
    - $0.1\text{m} / 343\text{m/s} = 0.29\text{ms}$
  - Propagation delay for aluminum
    - $0.1\text{m} / 6320\text{m/s} = 0.015\text{ms}$
  - Number of samples sent in 0.275ms
    - $0.275\text{ms} * 44100 \text{ samples / sec} = \mathbf{12 \text{ samples}}$

Generate the sound **N** times and gather  $12 * \mathbf{N}$  samples

# Considering only direct sound

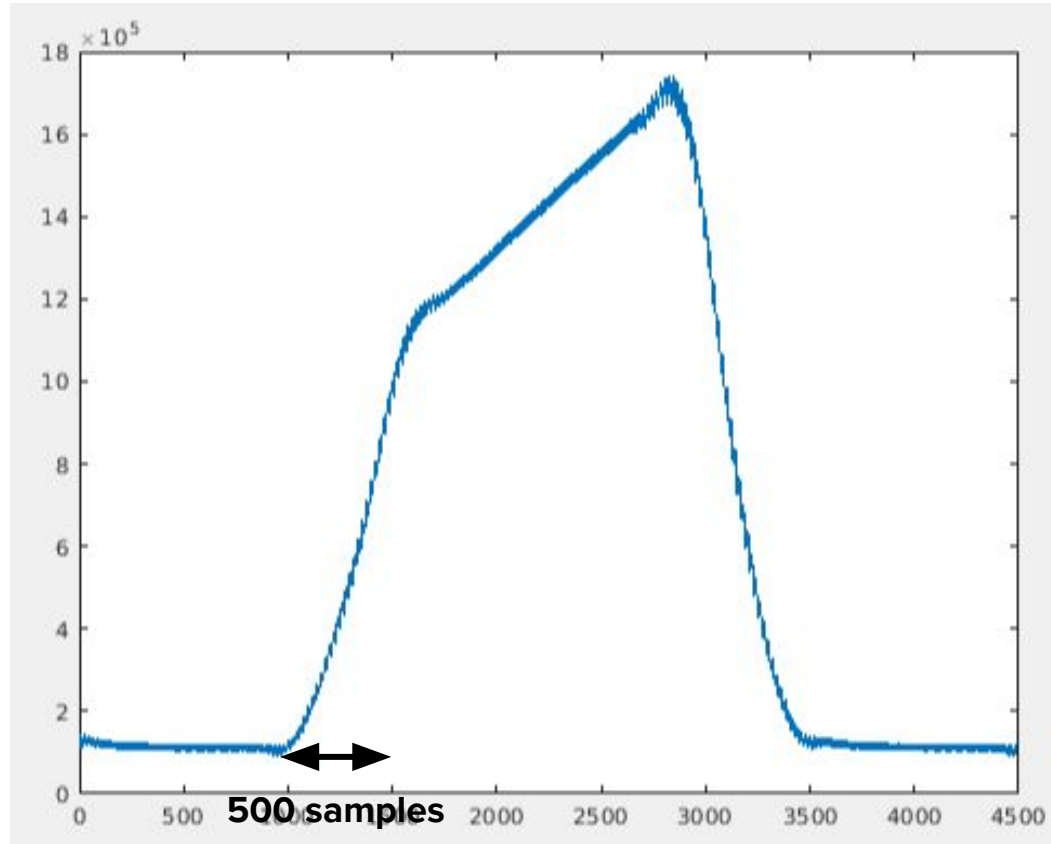
- Difference in speed of sound in air and solid

Medium (20 °C)	Speed of Sound Waves (m/s)
Dry Air	343
Water	1437
Wood	3850
Glass	4540
Aluminum	6320

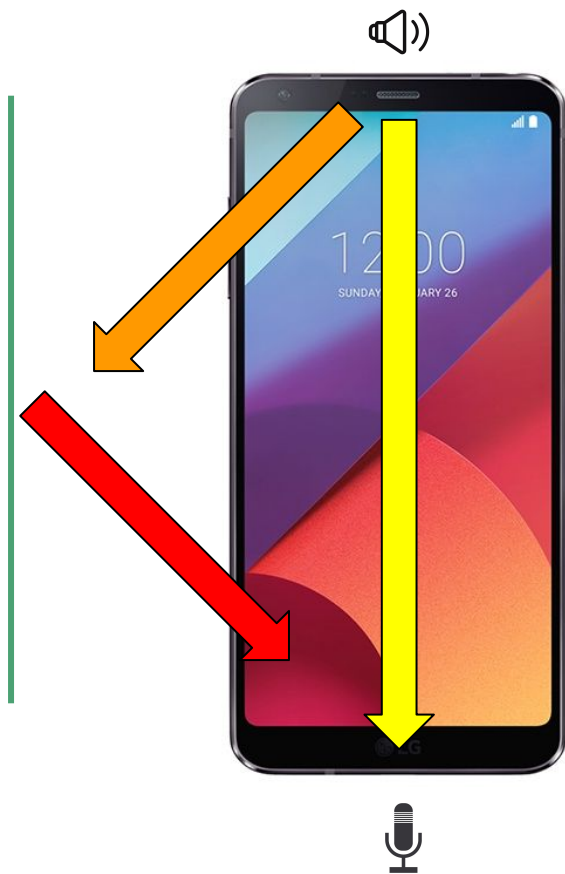
- Smartphone with aluminum, 10cm
  - Time for the first sample comes through air
    - $0.1\text{m} / 343\text{m/s} = 0.29\text{ms}$
  - Propagation delay for aluminum
    - $0.1\text{m} / 6320\text{m/s} = 0.015\text{ms}$
  - Number of samples sent in 0.275ms
    - $0.275\text{ms} * 44100 \text{ samples / sec} = \mathbf{12 \text{ samples}}$

Generate the sound  $N$  times and gather  $12*N$  samples

# Considering only direct sound



# Challenge in prior approach



Goal: detecting direct sound

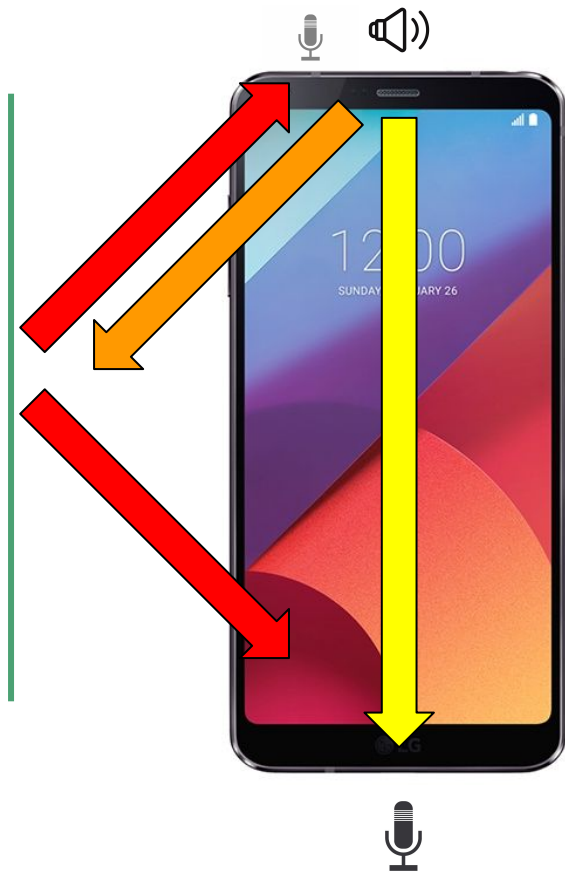


Approach 1: using very few samples first detected



Challenge: difficult to capture clear direct sound

# New approach: Canceling the reflected sound



Goal: detecting direct sound



Approach 1: using very few samples first detected

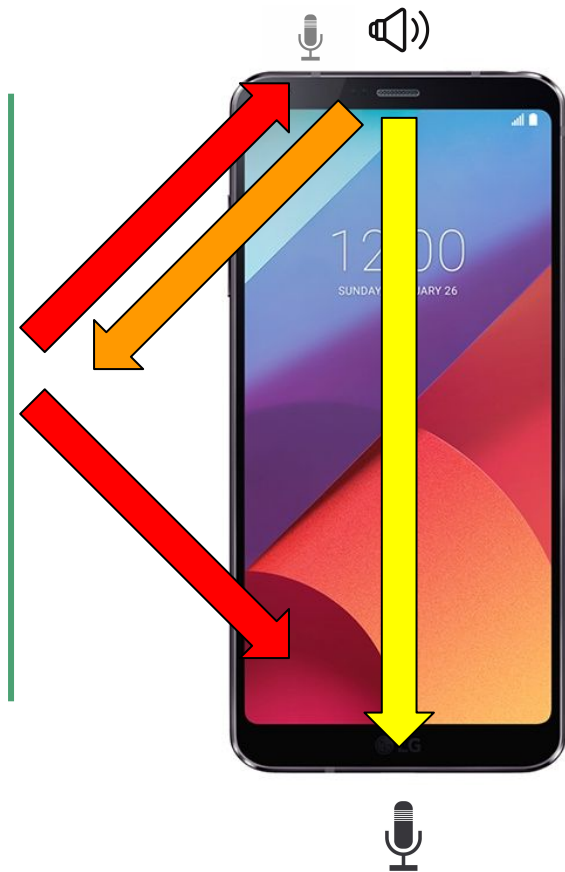


Challenge: difficult to capture clear direct sound




Approach 2: canceling reflected sound with another record

# New approach: Canceling the reflected sound



Common smartphone have 2 microphones, one on top and one on bottom

→ We can get **stereo records**

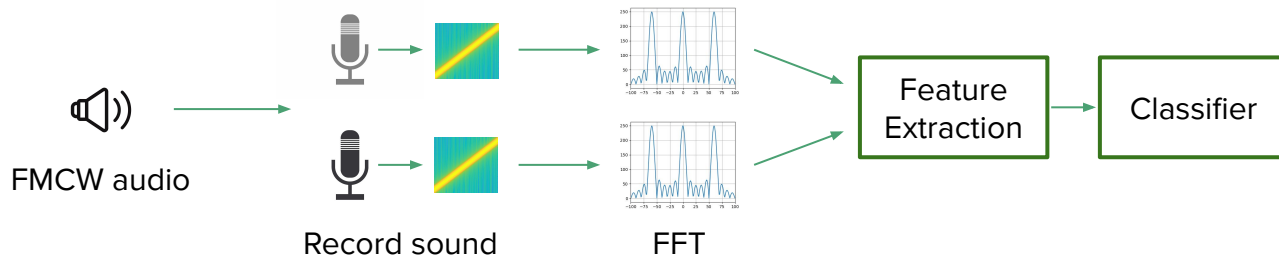
 Speaker (on top)

 Camcorder microphone (on top)

 Primary microphone (on bottom)

# Overall system of GraspTracker

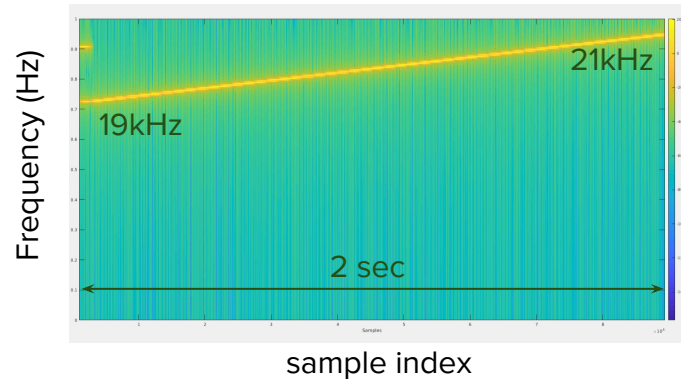
- Speaker (on top)
  - Original sound
- Camcorder microphone (on top)
  - Direct sound 1 + Reflected sound
- Primary microphone (on bottom)
  - Direct sound 2 + Reflected sound



# Evaluation

- FMCW audio structure

Spectrogram

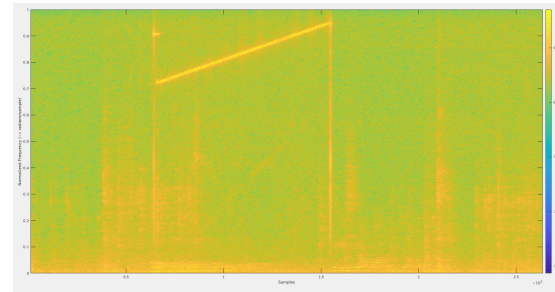
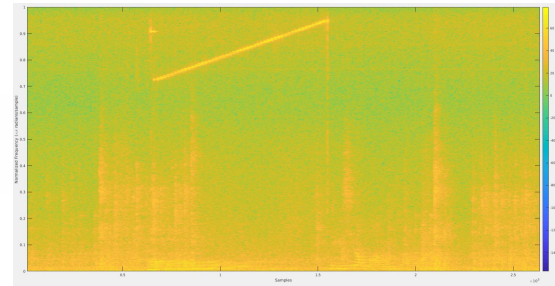
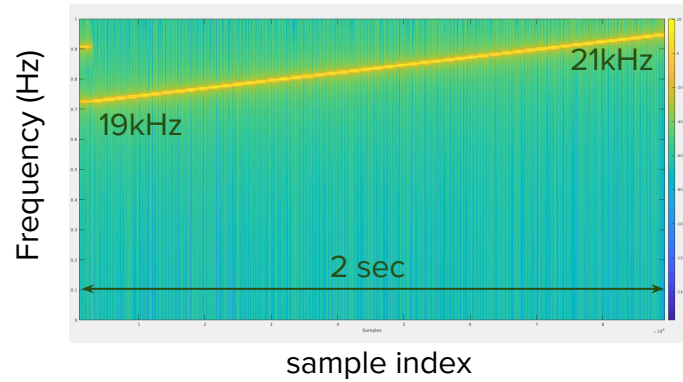




# Evaluation

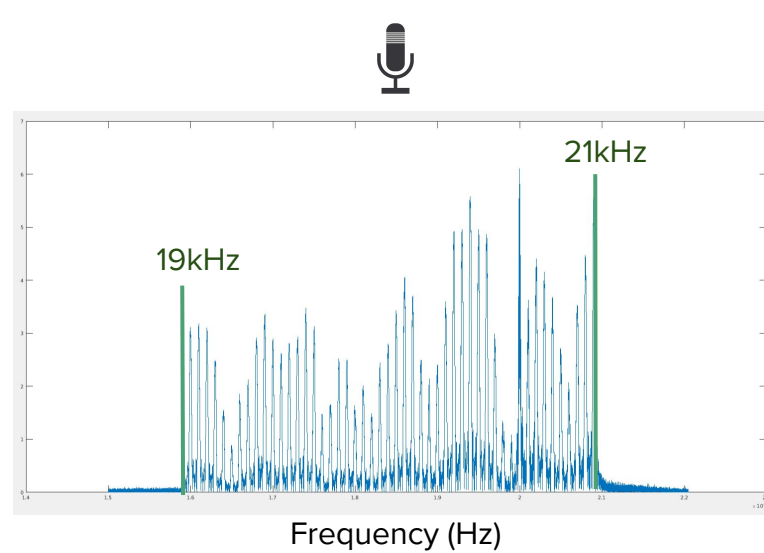
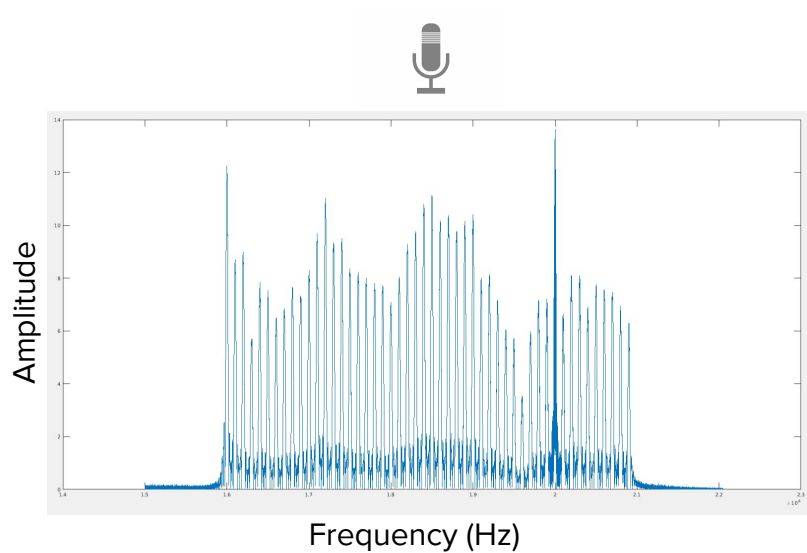
- FMCW audio structure

Spectrogram



# Evaluation

- FFT result



# Evaluation

- Feature extraction
  - From each FFT result, extracted 132 features
  - Each feature means the average amplitude of 50Hz window on FFT result
  - 264 features in total
- Label class
  - 7 classes
  - on\_table / one\_left\_hand / one\_right\_hand / two\_hands  
landscape\_left\_hand / landscape\_right\_hand / landscape\_two\_hands
- Classification
  - Dataset
    - 50 data for each class → 350 data
    - 40 training data, 10 training data for each class
  - Used SVM for classification
  - **Classification accuracy: 60%**

# GraspTracker

## Demo