Project L.A.K.E.

Logging of Acoustic Keyboard Emanations

Team A2: Ronit Banerjee, Kevin DeVincentis, James Zhang

Using Sound as a Keylogger

- Determine what a person is typing based on the sound of their keystrokes
- Exploit small differences in key sounds
- Ultimate goal: determine passwords from recordings of typing

Practicality

- Can be used in real world scenario
  - E.g. Libraries
- Custom low-power, small wireless sensor package.
- Learning process on a laptop done in 60 mins.
- Device has a long battery life, allowing us to collect more data

Requirements

Physical

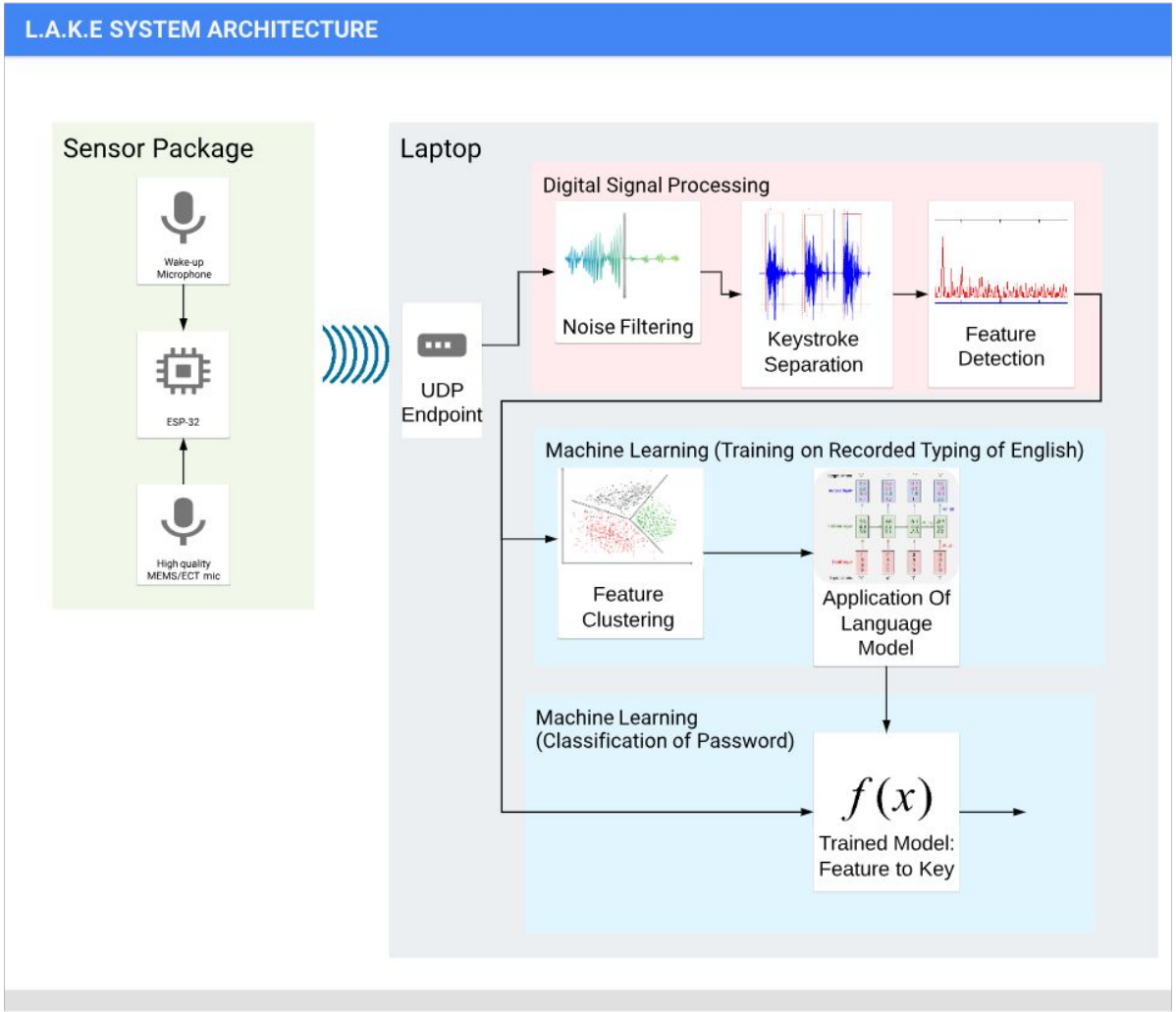
- Last for 24 hours (4 hours active) on a common 2000 mAh battery pack
- PCB size: 2 inches x 3 inches
- Device placed within 6 inches of keyboard

Computation

- Training data: 10 minutes of recorded English typing
- Training time: 60 minutes on modern laptop
- No English language model applied on output

Target accuracy

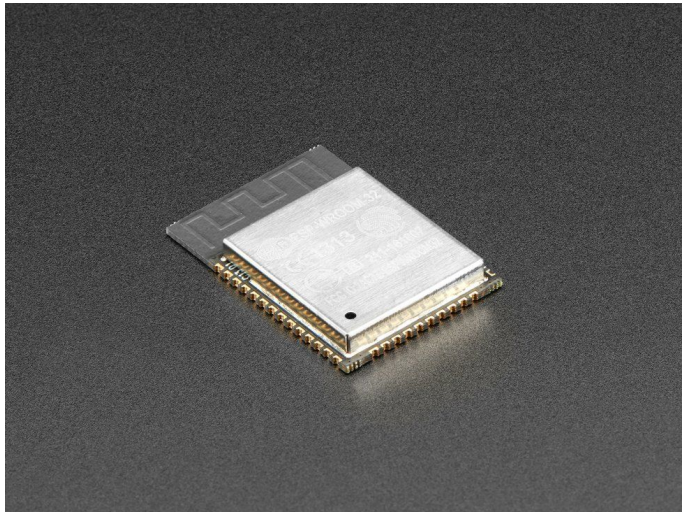
- 80% of 10-character passwords can be generated in fewer than 75 attempts
    - *Keyboard Acoustic Emanations Revisited* - Zhuang et al.



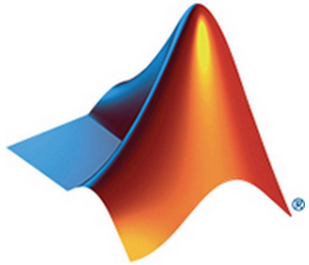
- Three Main Components
  - Embedded Sensor Package
  - Signal Processing
  - Machine Learning

Approach: Building a Sensor Package

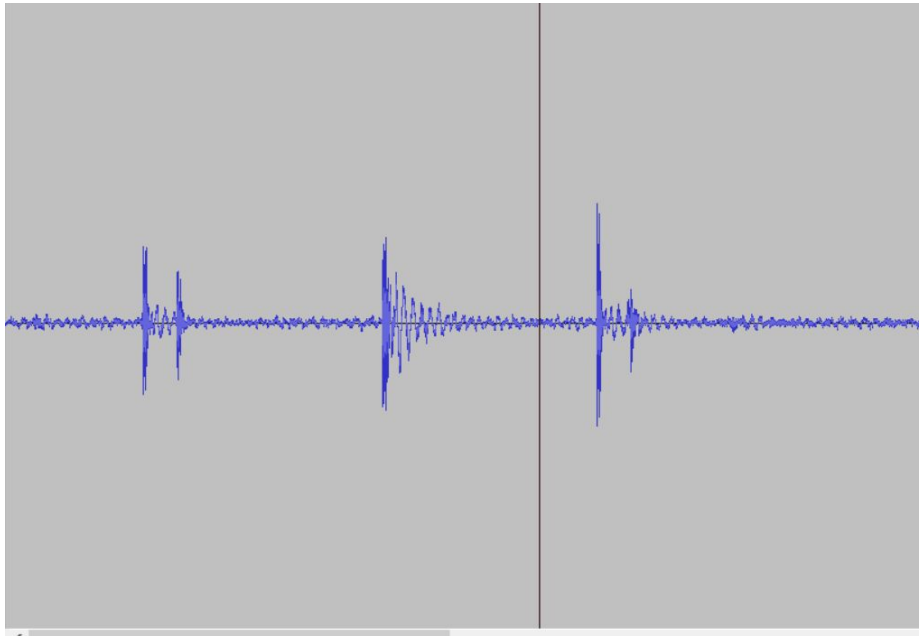
- Low-power sensor suite
- Use EC/MEMS microphone and accelerometer
- Custom hardware allows for more control over sensors
- ESP32 microcontroller for wireless communication
- Ultra low power wakeup microphone like the Vesper VM1010



Approach: Signal Processing

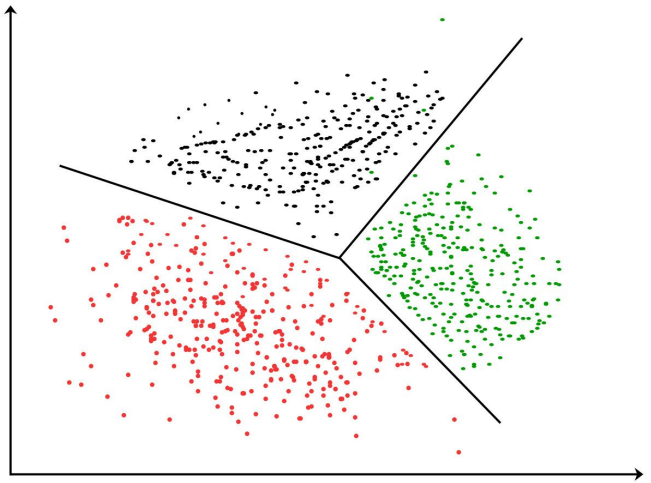


- Noise Reduction
- 3 distinct sounds of a keystroke
- Feature selection:
  - Frequency response
  - Time between keystrokes
  - Amplitude



Approach: Machine Learning Techniques

- Cluster keystrokes into different classes based on acoustic features
- Apply language model to match clusters to keys
  - English Language Statistics
  - HMM
- Apply cluster classification to guess typed letter
- For English text, apply language model again to improve accuracy



Challenges

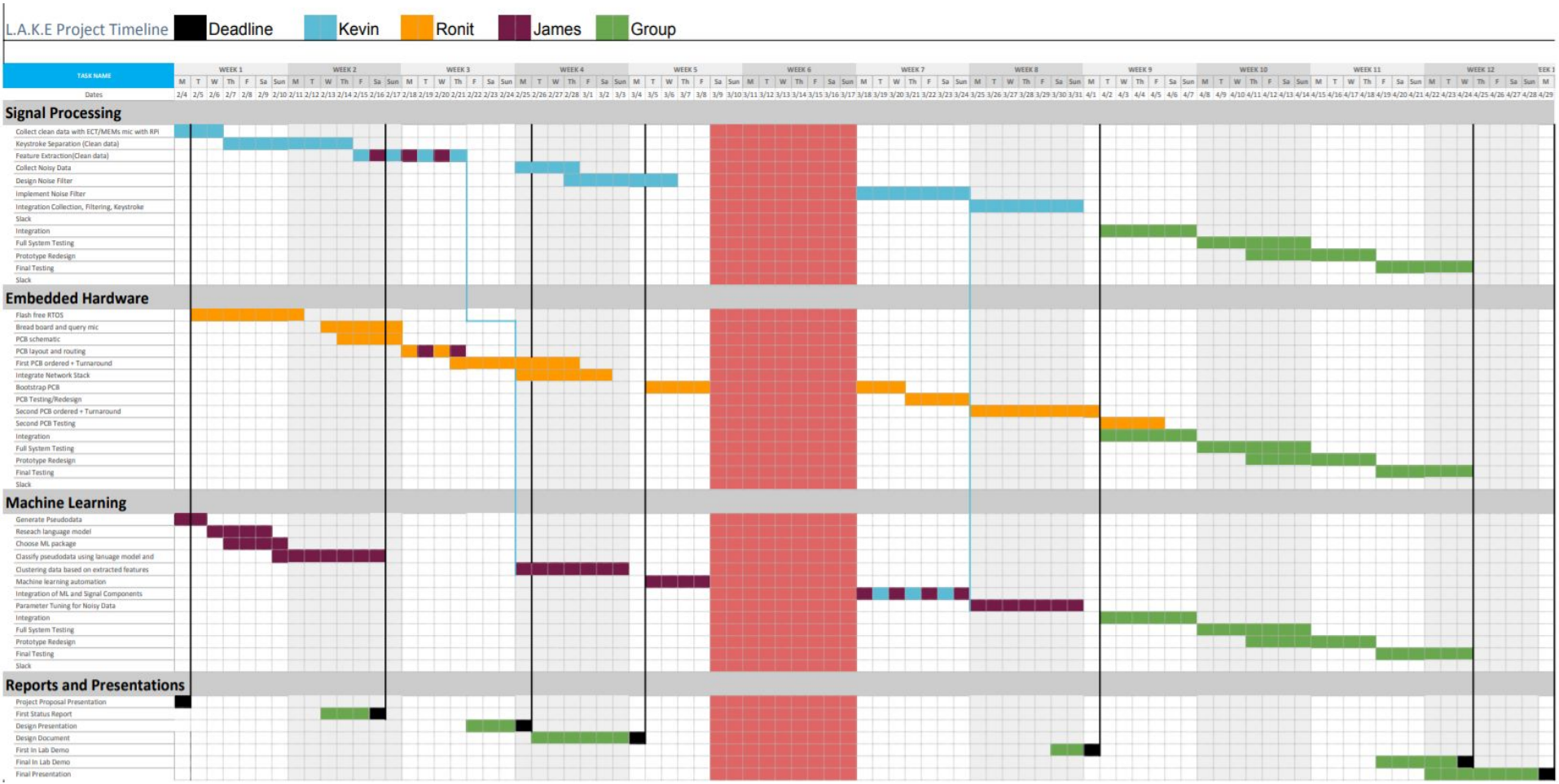
- Building a pcb in the form factor required
- Reducing noise in recorded audio
- Accurately seperating keystrokes without loss
- Determining features of keystrokes to learn on
- Learning less common letters like q, z, and x

Metrics and Testing

- Metric: Power consumption of sensor package
  Test:
  - Measure current consumption in active/sleep modes
  - Test: Stress test in real environment (HH1303) for 24 hours

- Metric: Accuracy
  - Individual character/word accuracy
  - Password accuracy (both 75 and 3 tries)

  Test:
  - Measure occurrence of misclassification in typed English text in test set
  - For each training set, generate a set of random 10-character passwords
  - Measure occurrence of correct password in top 75 guesses


L.A.K.E Project Timeline
Deadline
Kevin
Ronit
James
Group
TASK NAME
WEEK 1
WEEK 2
WEEK 3
WEEK 4
WEEK 5
WEEK 6
WEEK 7
WEEK 8
WEEK 9
WEEK 10
WEEK 11
WEEK 12
Dates
Signal Processing
Collect clean data with ECT/MEMs mic with RPi
Keystroke Separation (Clean data)
Feature Extraction(Clean data)
Collect Noisy Data
Design Noise Filter
Implement Noise Filter
Integration Collection, Filtering, Keystroke
Slack
Integration
Full System Testing
Prototype Redesign
Final Testing
Slack
Embedded Hardware
Flash free RTOS
Bread board and query mic
PCB schematic
PCB layout and routing
First PCB ordered + Turnaround
Integrate Network Stack
Bootstrap PCB
PCB Testing/Redesign
Second PCB ordered + Turnaround
Second PCB Testing
Integration
Full System Testing
Prototype Redesign
Final Testing
Slack
Machine Learning
Generate Pseudodata
Reseach language model
Choose ML package
Classify pseudodata using language model and
Clustering data based on extracted features
Machine learning automation
Integration of ML and Signal Components
Parameter Tuning for Noisy Data
Integration
Full System Testing
Prototype Redesign
Final Testing
Slack
Reports and Presentations
Project Proposal Presentation
First Status Report
Design Presentation
Design Document
First In Lab Demo
Final In Lab Demo
Final Presentation