Team B2: Jamming Attack on Voice Activated Systems

Final Presentation

Cyrus Daruwala, Eugene Luo, Spencer Yu

Recap

Motivation: Previous research has shown attacks to covertly manipulate voice activated systems are possible, and we are aiming to show the relevance of these attacks by demonstrating them on commodity hardware.

Goal: reduce the accuracy with which the wake word (“Hey Siri”) is recognized.

Solution:

1. Determine the input to jam the wake word (done)
2. Ensure that the jamming input is within latency bounds (done)
3. Reduce the % of false positives without compromising on the % of false negatives (goal for next week)

Changes to Solution Approach: MFCC + DTW

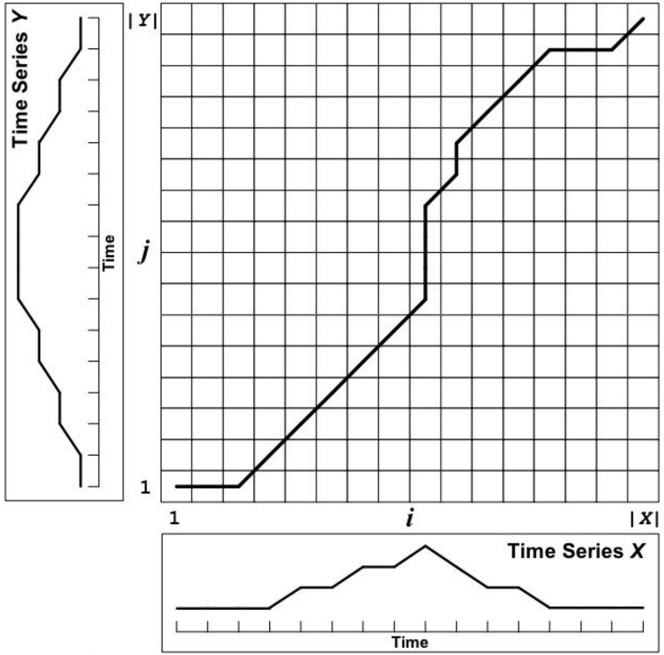


Figure 2. A cost matrix with the minimum-distance warp path traced through it.

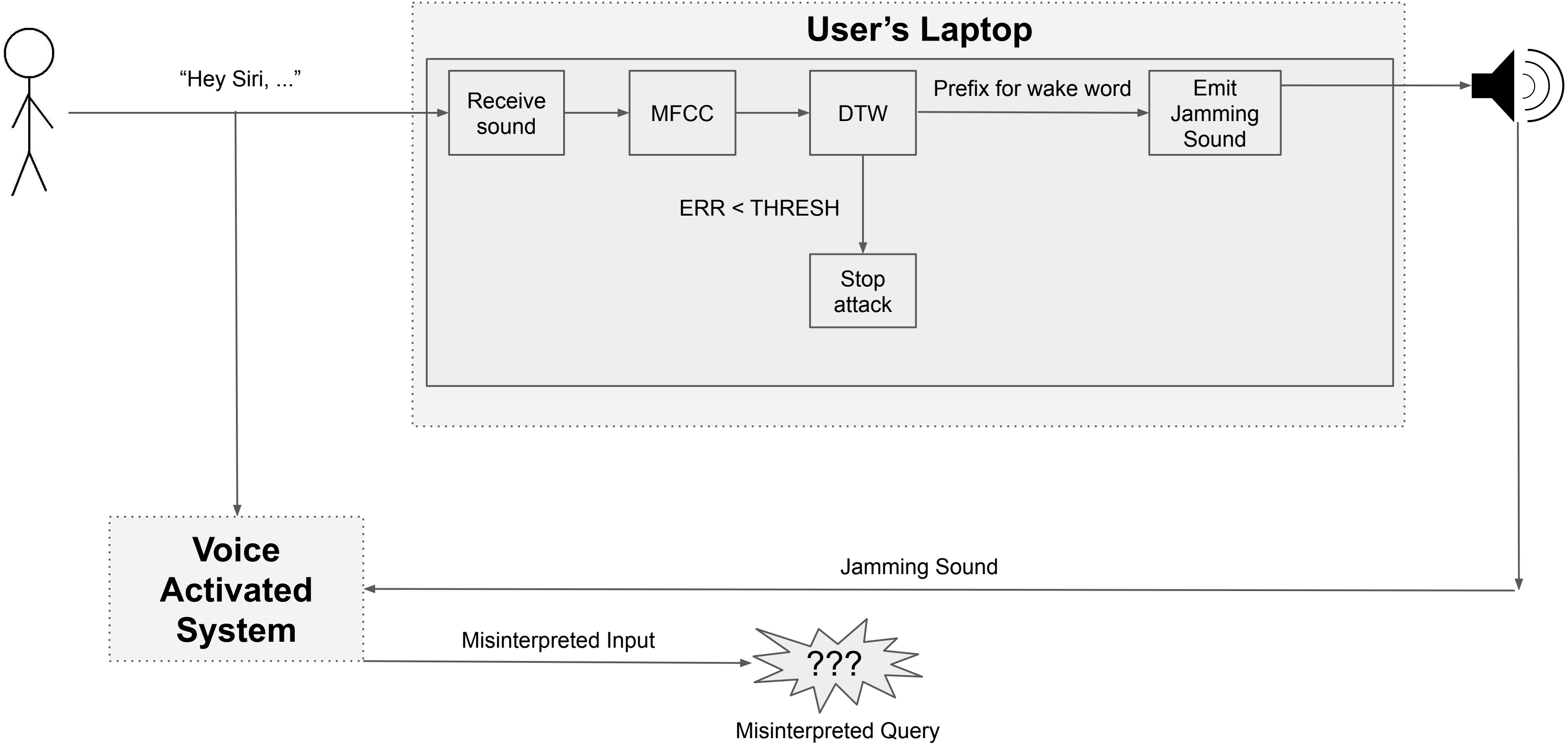
Source: https://cs.fit.edu/~pkc/papers/tdm04.pdf

Overall goal: Compare the similarity between two sounds of different lengths by aligning their identifying characteristics.

MFCC: coefficients that characterize the speech

DTW: matches characteristics in sound by stretching/compressing

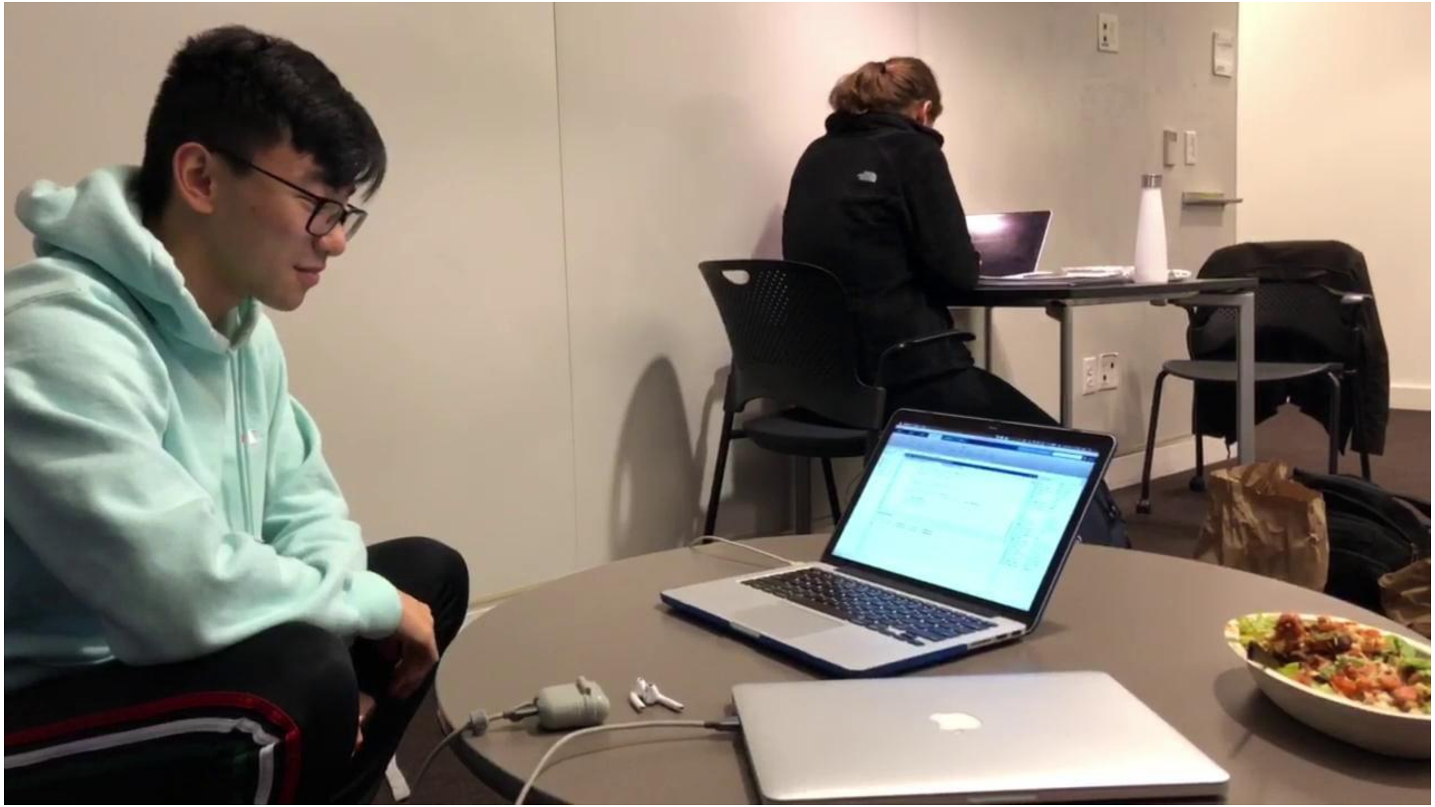
System Diagram



Demo: Complete Solution

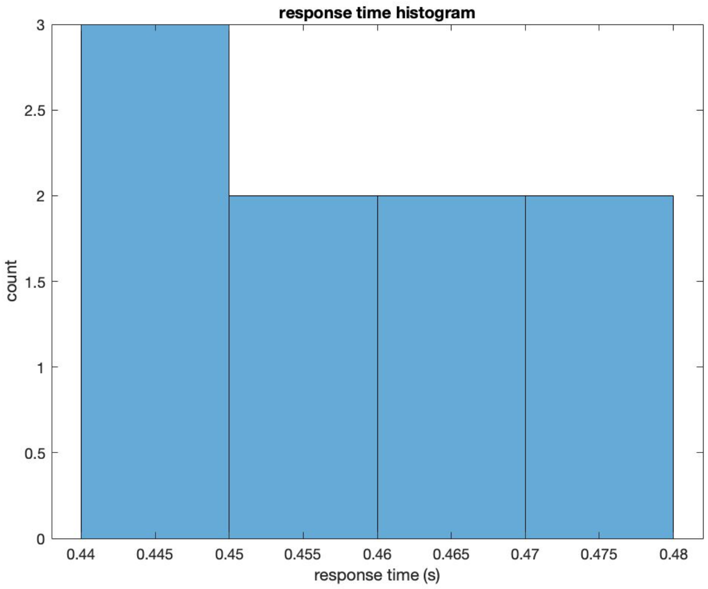
IMG_2782.mov

Demo: Selective Jamming



IMG_2786.mov

Measuring Performance



How do we verify the overall **performance and speed** of our system?

Latency: time sync + manual validation via MATLAB scripts

- Response time average **462.5ms** over 10 queries

System Validation: tested against Spencer’s iPhone 7 Plus

- Response rate drops to **46.67%** over 30 queries

Metrics by Sound

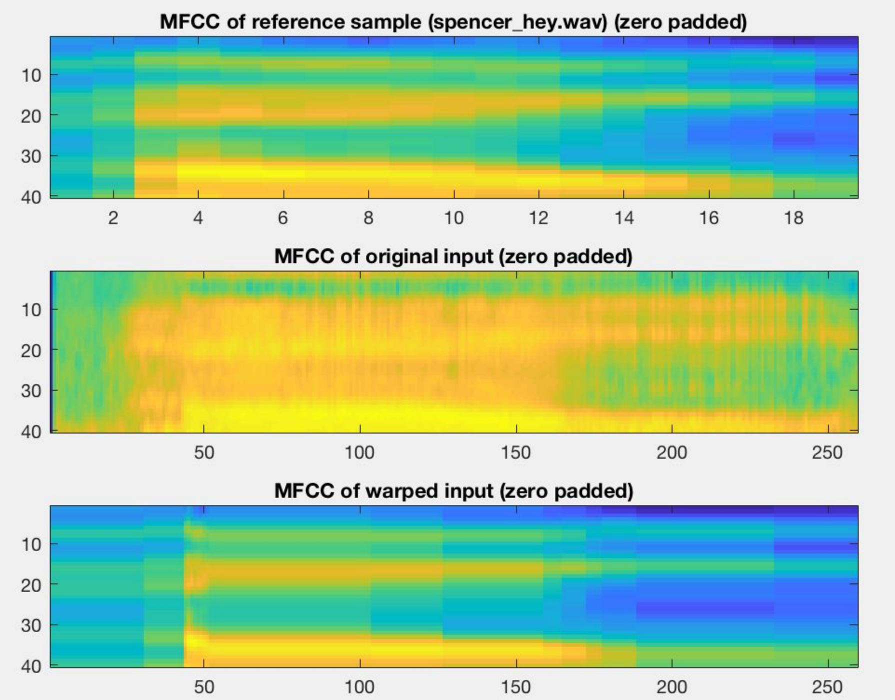
Noise Type	White Noise	Music w/ vocals[1]	Music w/o vocals[2]	Podcasts[3]	Human Speech[4]
False Negative Rate	N/A	N/A	N/A	0%	0%
False Positive Rate	0%	3.43% 20.3/min	2.105% 15/min	2.7% 17.5/min	0.263%, 5.22/min

False Negative Rate: % where attack does not trigger when it should

False Positive Rate: % where attack triggers when it should not

[1] - Sound of a box fan (https://www.youtube.com/watch?v=Qq_wJFvhfrg)
[1] - “Hey Jude” by the Beatles
[2] - “The Sleeping Beauty, Op. 66, TH 13: No. 6, Valse” by Pyotr Ilyich Tchaikovsky, conducted by Eugene Ormandy
[3] - “Great Bitter Lake Association”, *99% Invisible* by Roman Mars
[4] - Spencer Yu

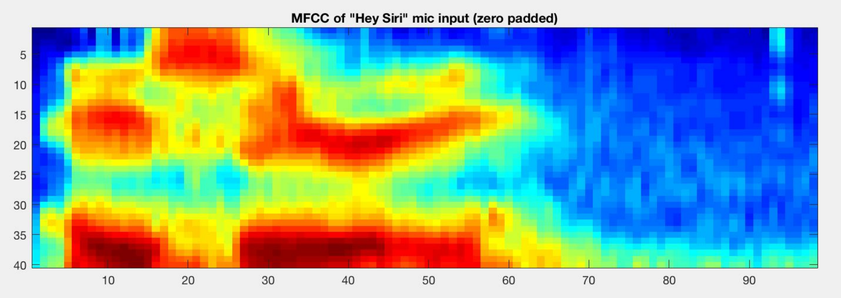
Validating Signal Processing



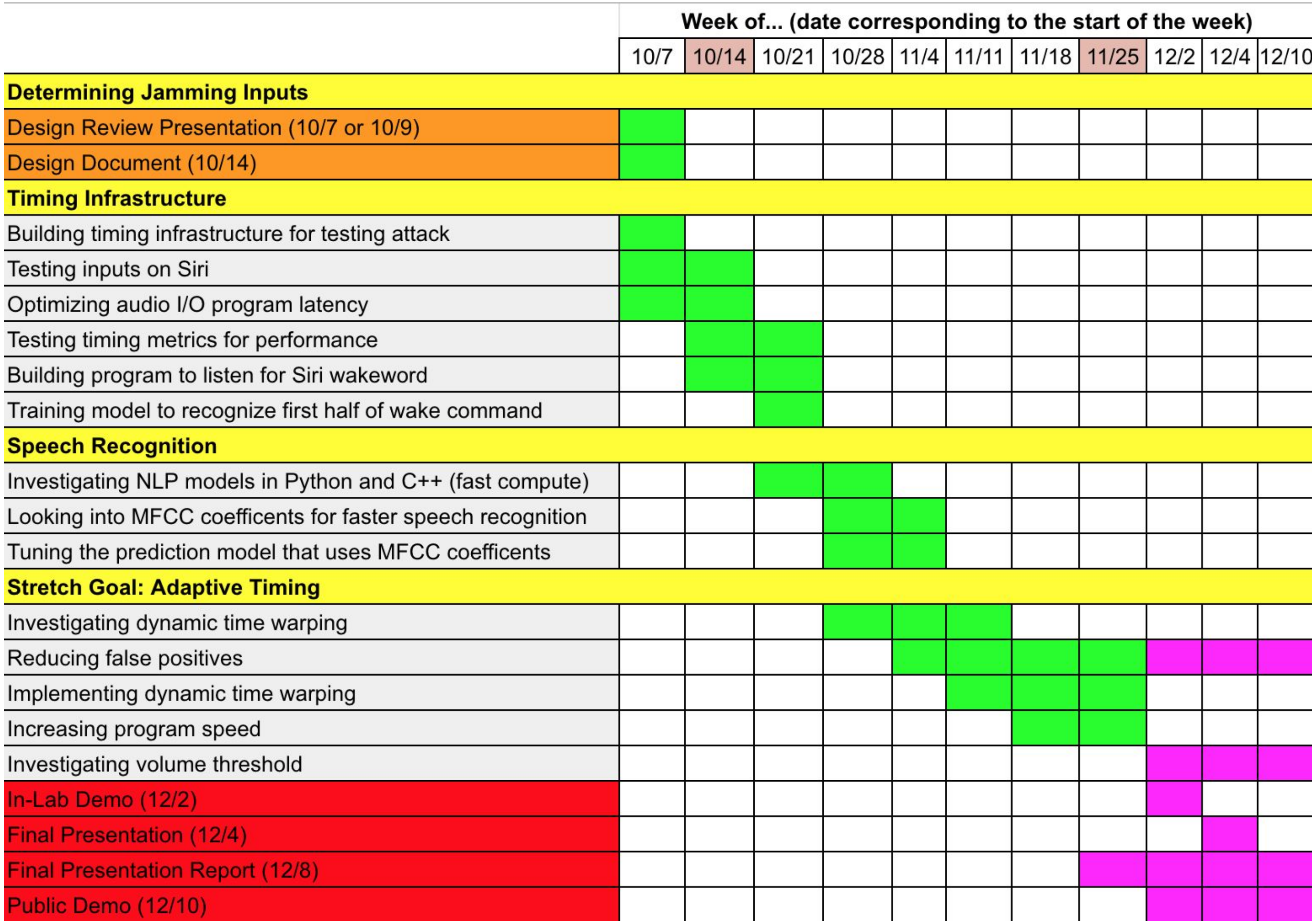
How do we verify each **individual component** of our attack works as expected?

MFCC: IDCT + spectrogram

DTW: IDCT + spectrogram
Play out sounds



Project Management



	Week of... (date corresponding to the start of the week)										
	10/7	10/14	10/21	10/28	11/4	11/11	11/18	11/25	12/2	12/4	12/10
Determining Jamming Inputs											
Design Review Presentation (10/7 or 10/9)	🟩										
Design Document (10/14)	🟩										
Timing Infrastructure											
Building timing infrastructure for testing attack	🟩										
Testing inputs on Siri	🟩	🟩									
Optimizing audio I/O program latency	🟩	🟩									
Testing timing metrics for performance		🟩	🟩								
Building program to listen for Siri wakeword		🟩	🟩								
Training model to recognize first half of wake command			🟩								
Speech Recognition											
Investigating NLP models in Python and C++ (fast compute)			🟩	🟩							
Looking into MFCC coefficents for faster speech recognition				🟩	🟩						
Tuning the prediction model that uses MFCC coefficents				🟩	🟩						
Stretch Goal: Adaptive Timing											
Investigating dynamic time warping				🟩	🟩	🟩					
Reducing false positives					🟩	🟩	🟩	🟩	🟪	🟪	🟪
Implementing dynamic time warping						🟩	🟩	🟩			
Increasing program speed							🟩	🟩			
Investigating volume threshold									🟪	🟪	🟪
In-Lab Demo (12/2)									🟪		
Final Presentation (12/4)										🟪	
Final Presentation Report (12/8)								🟪	🟪	🟪	🟪
Public Demo (12/10)									🟪	🟪	🟪

Obstacles Encountered

Choice of language

- **Before: Python**
- **After: Matlab**

Area of ECE Explored in Project

- **Before:** Software with signal support
- **After:** Largely signal processing with software implementation

Solution Approach

- **Before:** ML for generating/triggering adversarial input
- **After:** Signal processing

Lessons Learned

Scope of project

Having backup plans

Constant feedback from experts

Choose interesting projects

Take risks and try something new