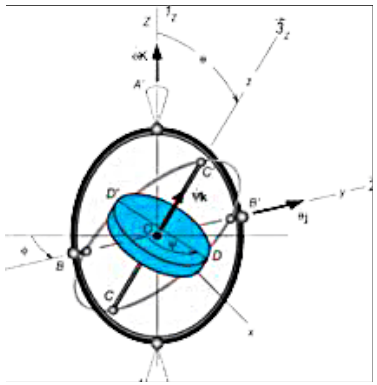
GYROPHONE

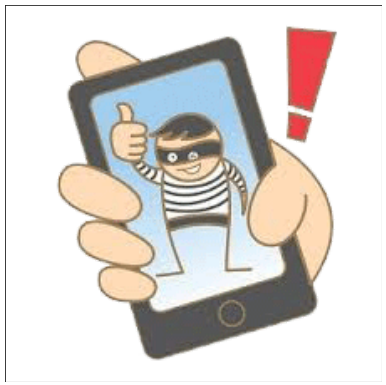
RECOGNIZING SPEECH FROM GYROSCOPE SIGNALS

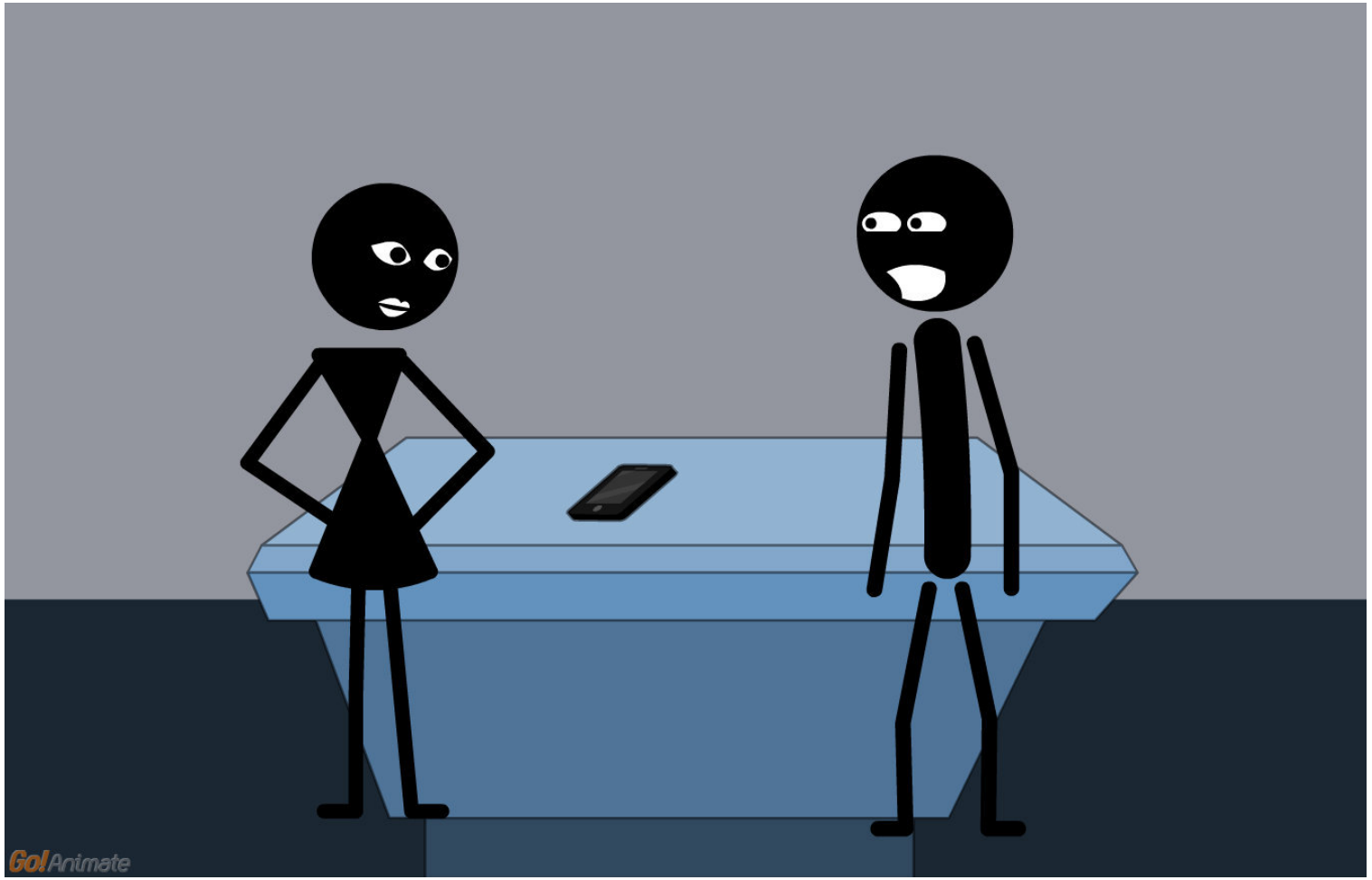
Yan Michalevsky [1], Gabi Nakibly [2] and Dan Boneh [1]

[1] Stanford University

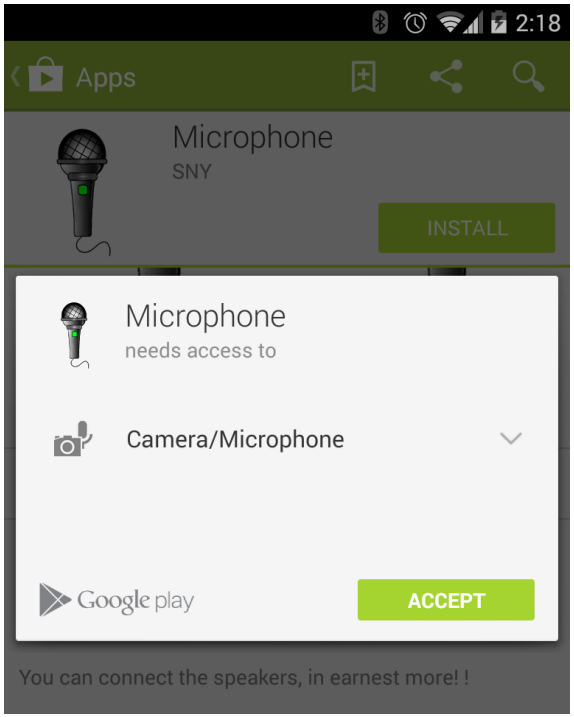
[2] National Research and Simulation Center, Rafael Ltd.





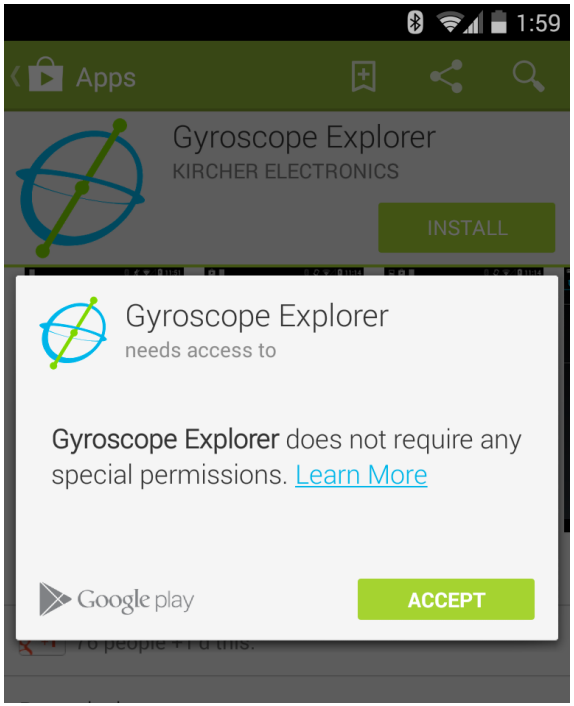

GoAnimate

MICROPHONE ACCESS



REQUIRES PERMISSIONS

GYROSCOPE ACCESS

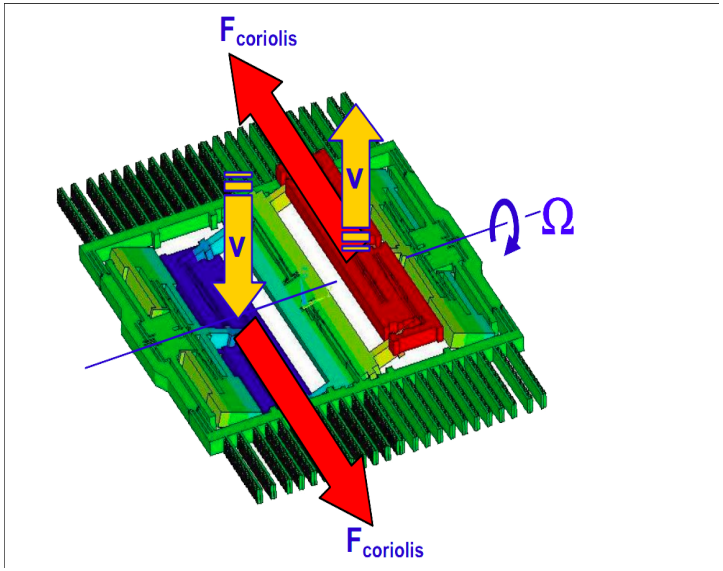


DOES NOT REQUIRE PERMISSIONS

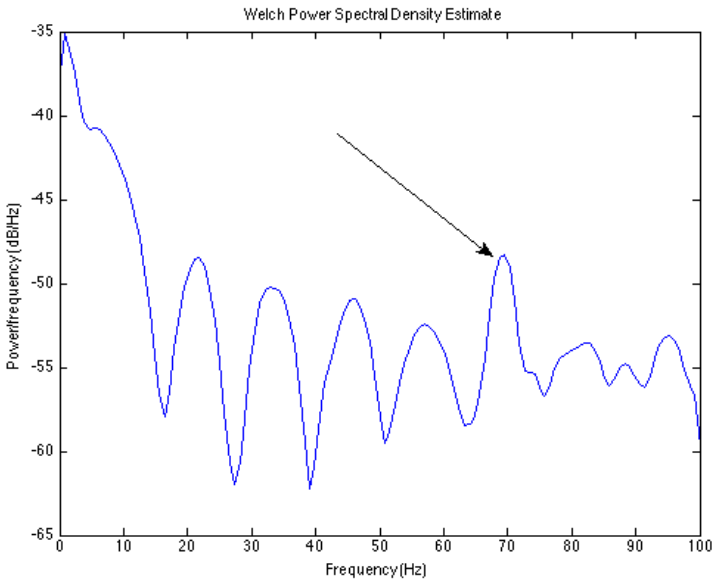
MEMS GYROSCOPES

Major vendors:

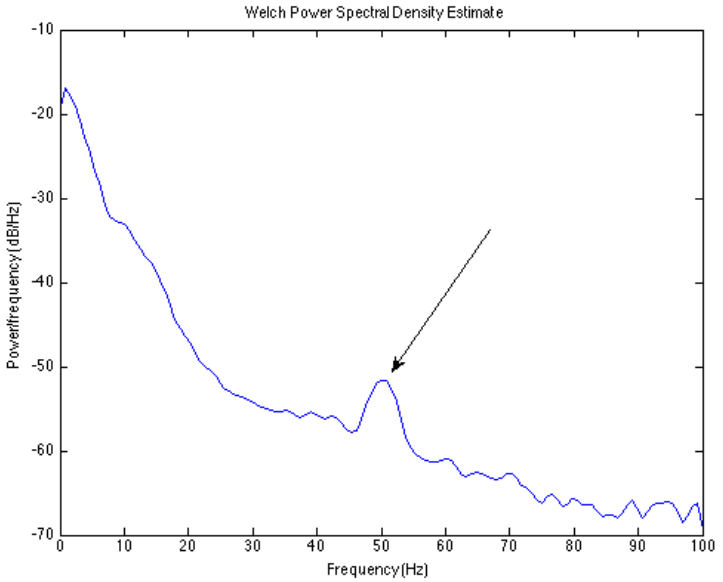
- STM Microelectronics (Samsung Galaxy)
- InvenSense (Google Nexus)



GYROSCOPES ARE SUSCEPTIBLE TO SOUND



70 HZ TONE POWER SPECTRAL DENSITY



50 HZ TONE POWER SPECTRAL DENSITY

GYROSCOPES ARE (LOUSY, BUT STILL) MICROPHONES

- Hardware sampling frequency:
  - InvenSense: up to 8000 Hz
  - STM Microelectronics: 800 Hz
- Software sampling frequency:
  - Android: 200 Hz
  - iOS: 100 Hz

GYROSCOPES ARE (LOUSY, BUT STILL) MICROPHONES

- Very low SNR (Signal-to-Noise Ratio)
- Acoustic sensitivity threshold: ~70 dB

  Comparable to a loud conversation.
- Sensitive to sound angle of arrival
- Directional microphone (due to 3 axes)

BROWSERS ALLOW GYROSCOPE ACCESS TOO

		Sampling Freq. [Hz]
Android 4.4	application	200
	Chrome	25
	Firefox	200
	Opera	20
iOS 7	application	100
	Safari	20
	Chrome	20

BROWSERS ALLOW GYROSCOPE ACCESS TOO

		Sampling Freq. [Hz]
Android 4.4	application	200
	Chrome	25
	Firefox	200
	Opera	20
iOS 7	application	100
	Safari	20
	Chrome	20

BROWSERS ALLOW GYROSCOPE ACCESS TOO

		Sampling Freq. [Hz]
Android 4.4	application	200
	Chrome	25
	Firefox	200
	Opera	20
iOS 7	application	100
	Safari	20
	Chrome	20

WebKit

BROWSERS ALLOW GYROSCOPE ACCESS TOO

		Sampling Freq. [Hz]
Android 4.4	application	200
	Chrome	25
	Firefox	200
	Opera	20
iOS 7	application	100
	Safari	20
	Chrome	20

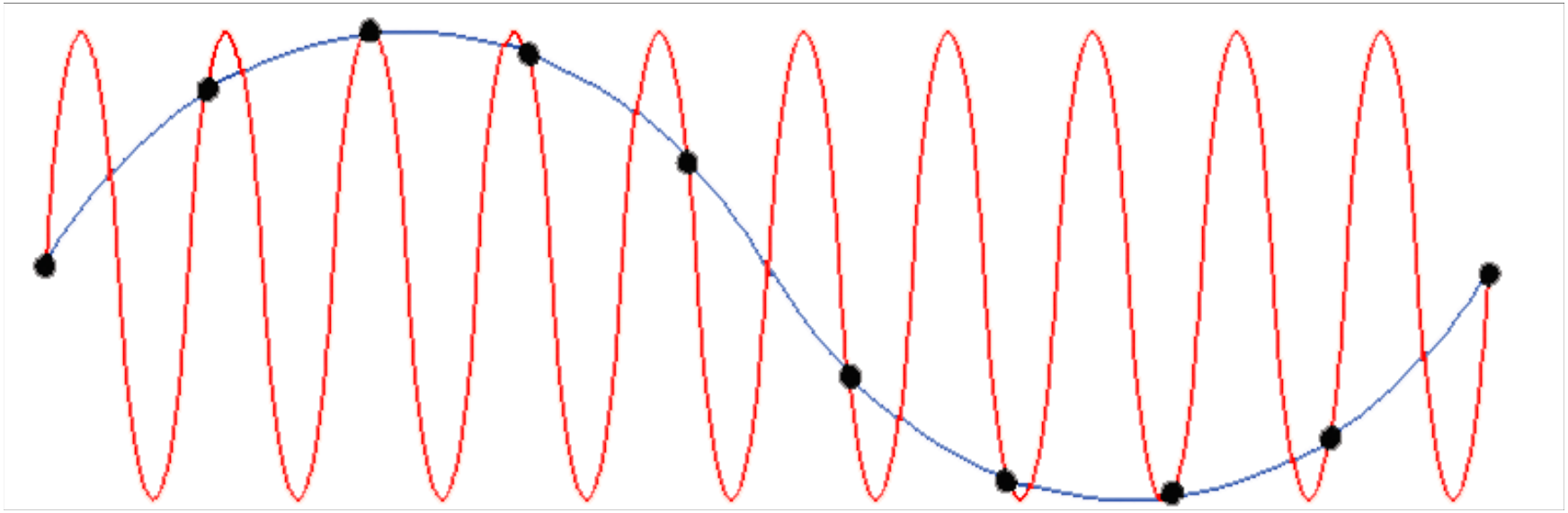
Gecko

PROBLEM: HOW DO WE LOOK INTO HIGHER FREQUENCIES?

SPEECH RANGE

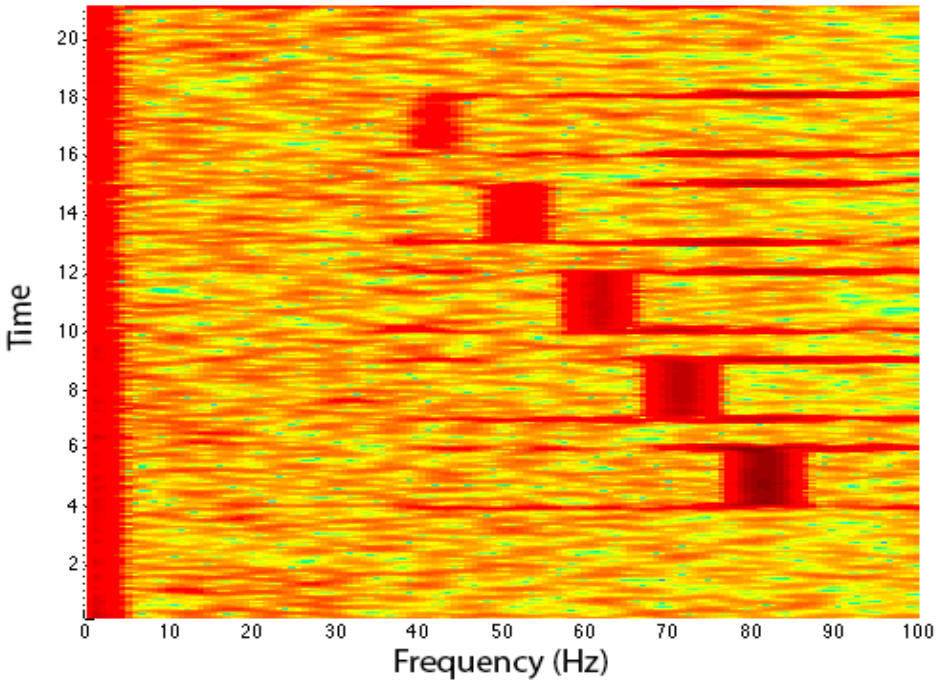
Adult male 85 - 180 Hz
Adult female 165 - 255 Hz

ALIASING



WE CAN SENSE HIGH FREQUENCY SIGNALS

DUE TO ALIASING



THE RESULT OF RECORDING TONES BETWEEN 120 AND 160 HZ ON A NEXUS 7 DEVICE

EXPERIMENTAL SETUP

- Room. Simple speakers.

  Smartphone.
- Subset of TIDigits speech

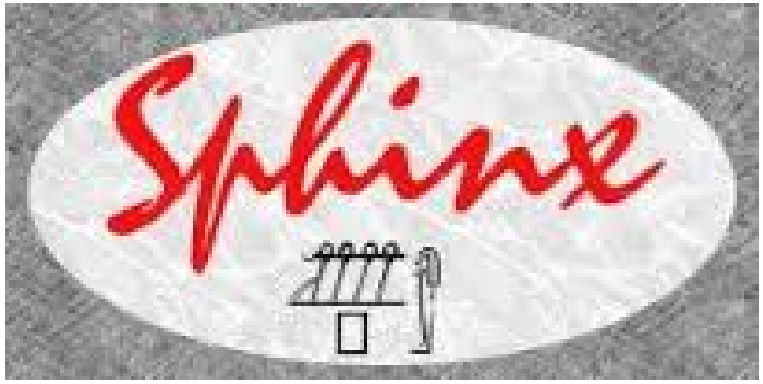
  recognition corpus
- 10 speakers × 11 samples × 2

  pronunciations = 220 total

  samples

SPEECH ANALYSIS USING A SINGLE GYROSCOPE

- Gender identification
- Speaker identification
- Isolated word recognition



- Speech recognition engine developed at CMU
- Tested for isolated word recognition
- 14% success rate (random guess is 9%)

PREPROCESSING

- All samples are converted to audio files in WAV format
- Upsampled to 8 KHz
- Silence removal

  (based on voiced/unvoiced segment classification)

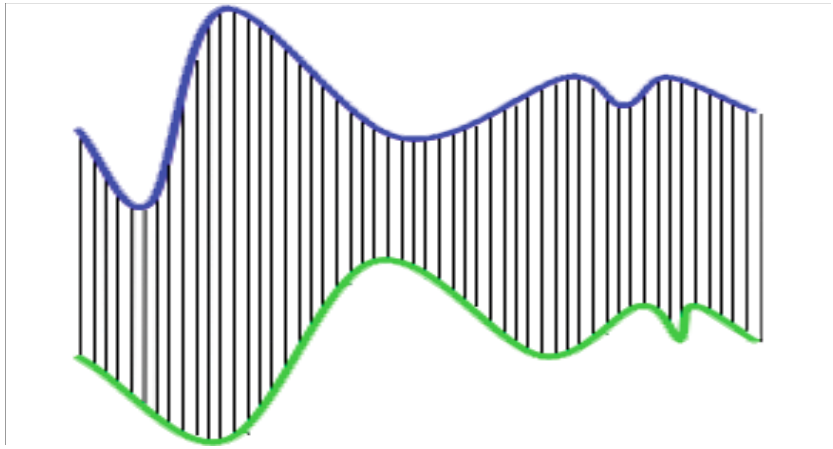
FEATURES

- MFCC - Mel-Frequency Cepstral Coefficients
  - Statistical features are used (mean and variance)
  - delta-MFCC
- Spectral centroid
- RMS energy
- STFT - Short-Time Fourier Transform

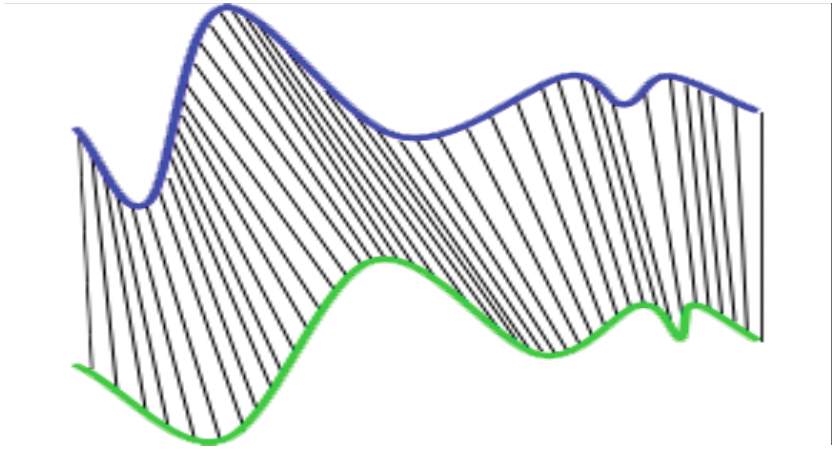
CLASSIFIERS

- SVM (and Multi-class SVM)
- GMM (Gaussian Mixture Model)
- DTW (Dynamic Time Warping)

DYNAMIC TIME WARPING



EUCLIDEAN DISTANCE



DTW DISTANCE

GENDER IDENTIFICATION

- Binary SVM with spectral features
- DTW with STFT features
  - Window size: 512 samples - corresponds to 64 ms under 8 KHz sampling rate

WE CAN SUCCESSFULLY IDENTIFY GENDER

NEXUS 4	84% (DTW)
GALAXY S III	82% (SVM)

Random guess probability is 50%

SPEAKER IDENTIFICATION

- Multi-class SVM and GMM with spectral features
- DTW with STFT features (same as before)

A GOOD CHANCE TO IDENTIFY THE SPEAKER

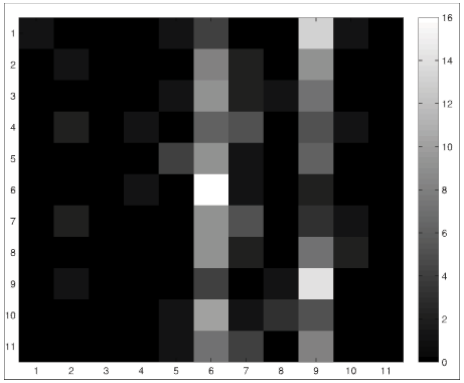
Nexus 4	Mixed Female/Male	50% (DTW)
	Female speakers	45% (DTW)
	Male speakers	65% (DTW)

Random guess probability is 20% for one gender and 10% for a mixed set

ISOLATED WORDS RECOGNITION

SPEAKER INDEPENDENT

Nexus 4	Mixed Female/Male	17% (DTW)
	Female speakers	26% (DTW)
	Male speakers	23% (DTW)



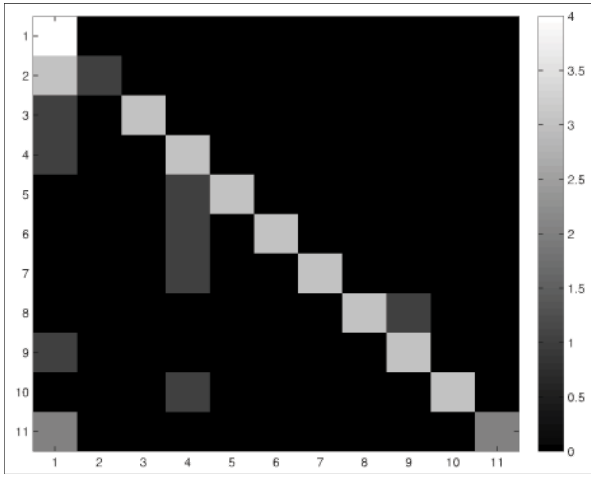
Confusion matrix corresponds to the mixed set results using DTW

Random guess probability is 9%

ISOLATED WORDS RECOGNITION

SPEAKER DEPENDENT

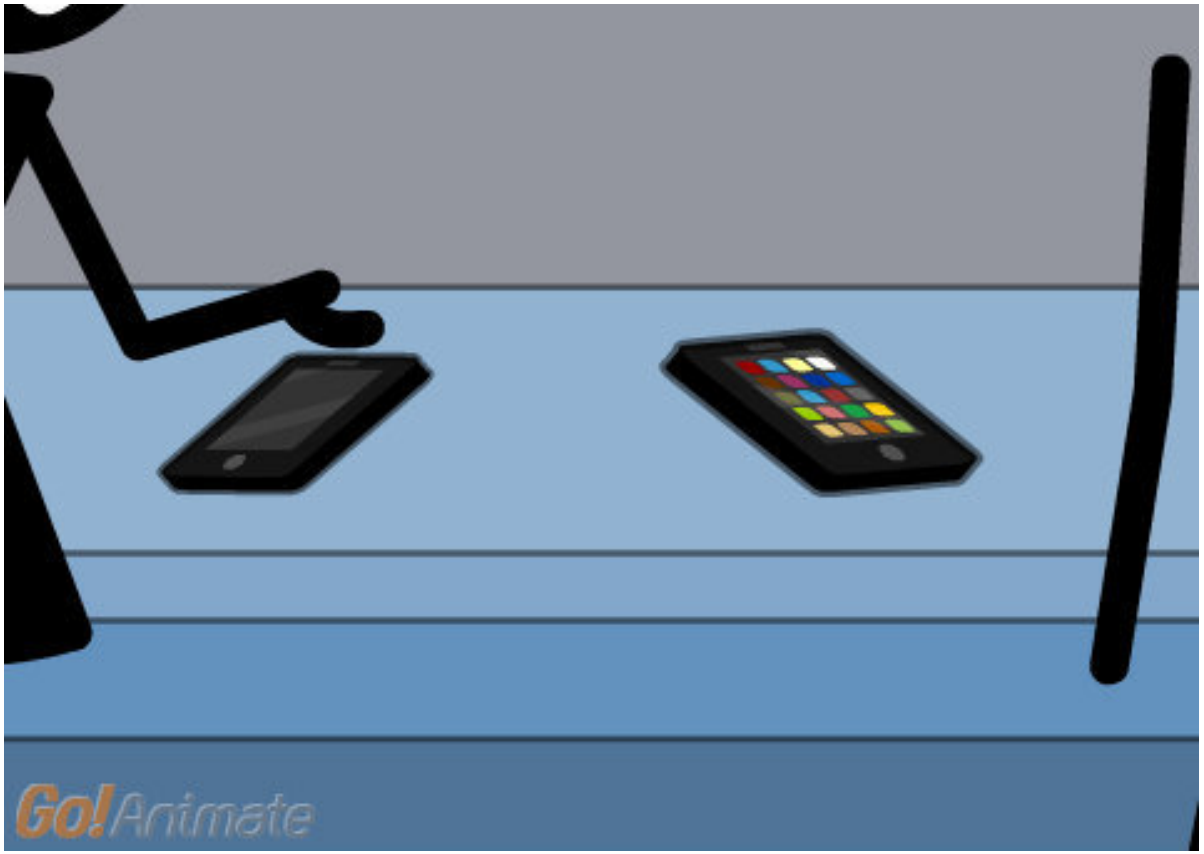
SVM	GMM	DTW
15%	5%	65%



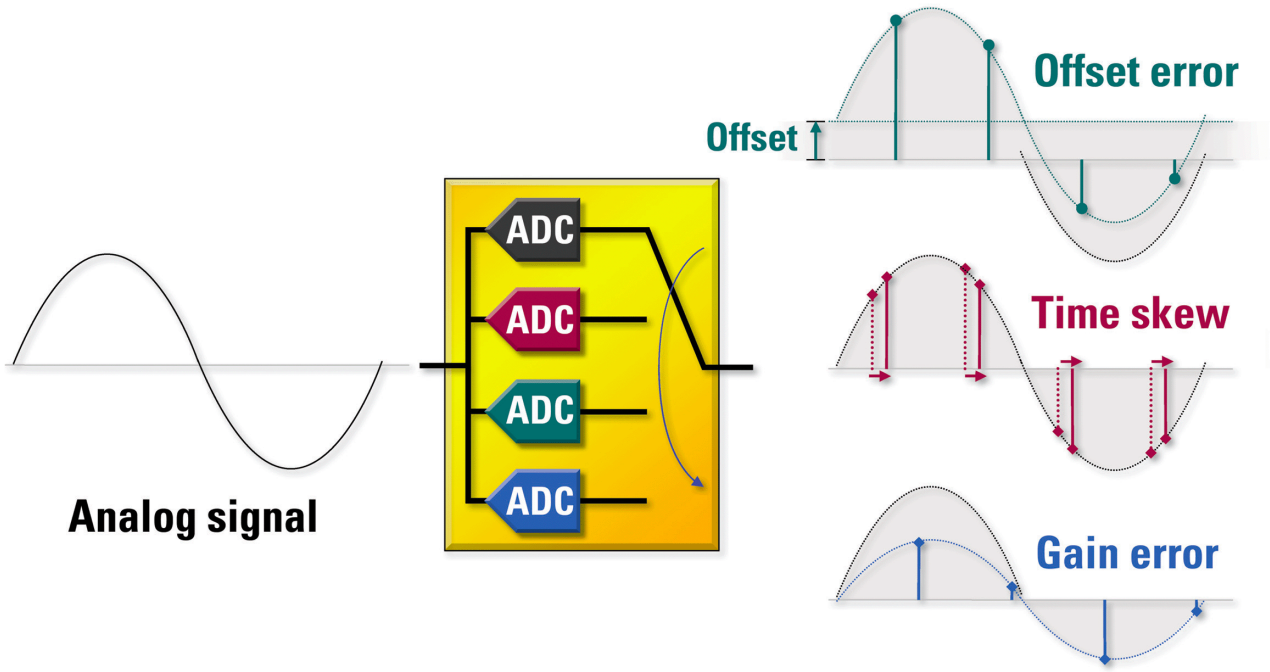
Confusion matrix corresponds to the DTW results

Random guess probability is 9%

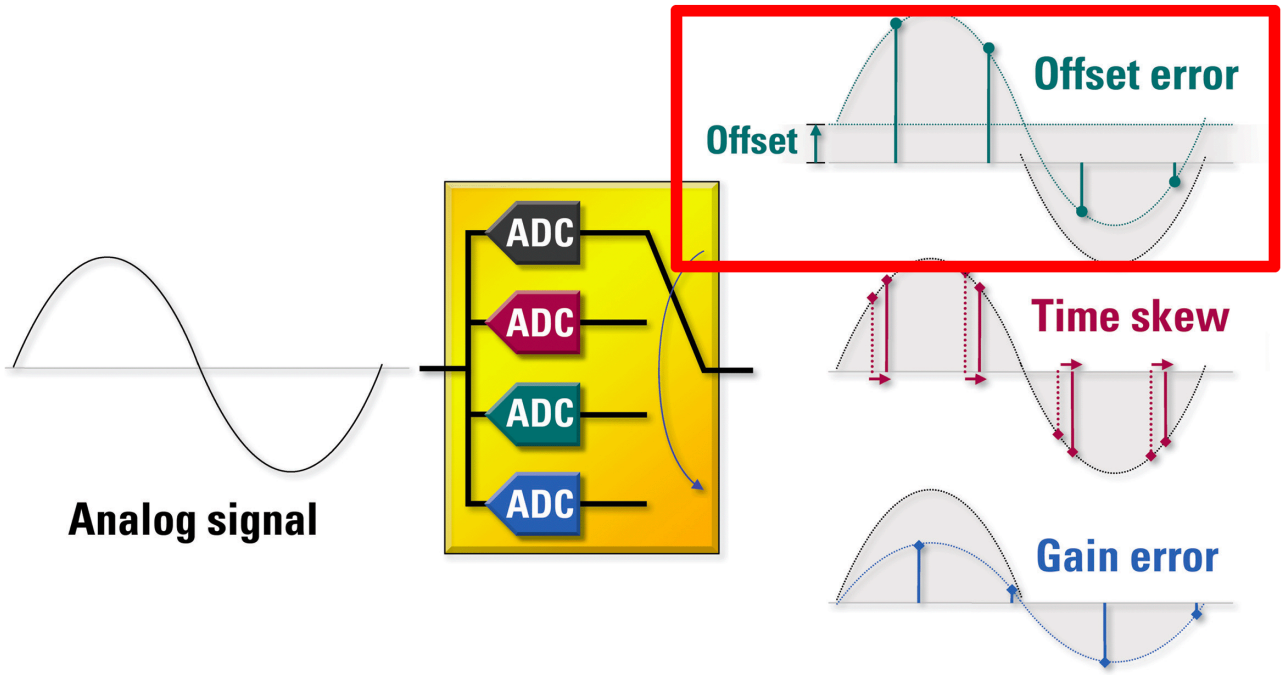
HOW CAN WE LEVERAGE EAVESDROPPING SIMULTANEOUSLY ON TWO DEVICES?



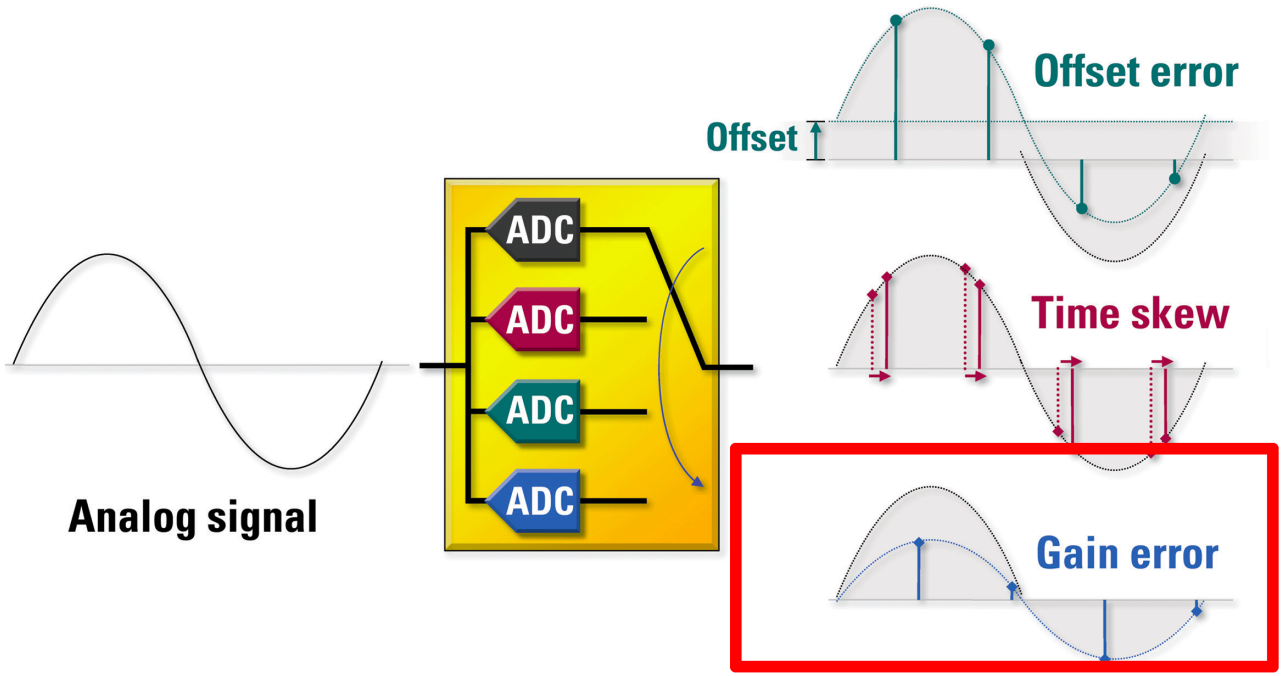
SIMILAR TO TIME-INTERLEAVED ADC's



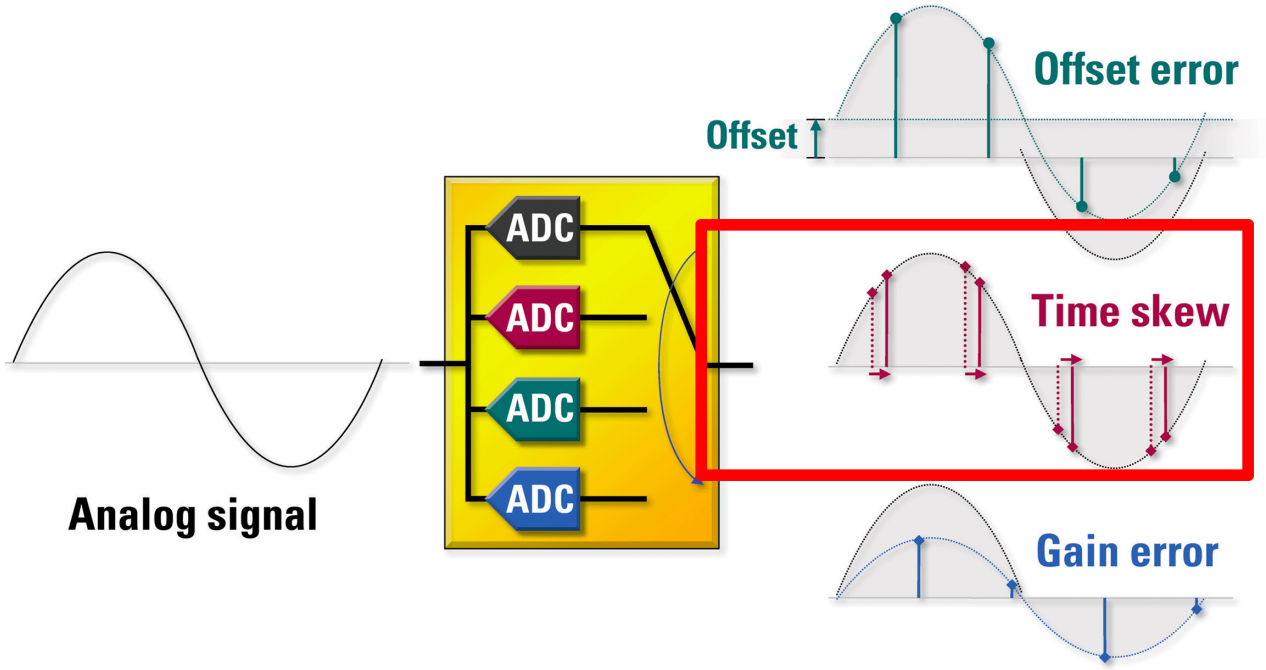
SIMILAR TO TIME-INTERLEAVED ADC's



SIMILAR TO TIME-INTERLEAVED ADC's

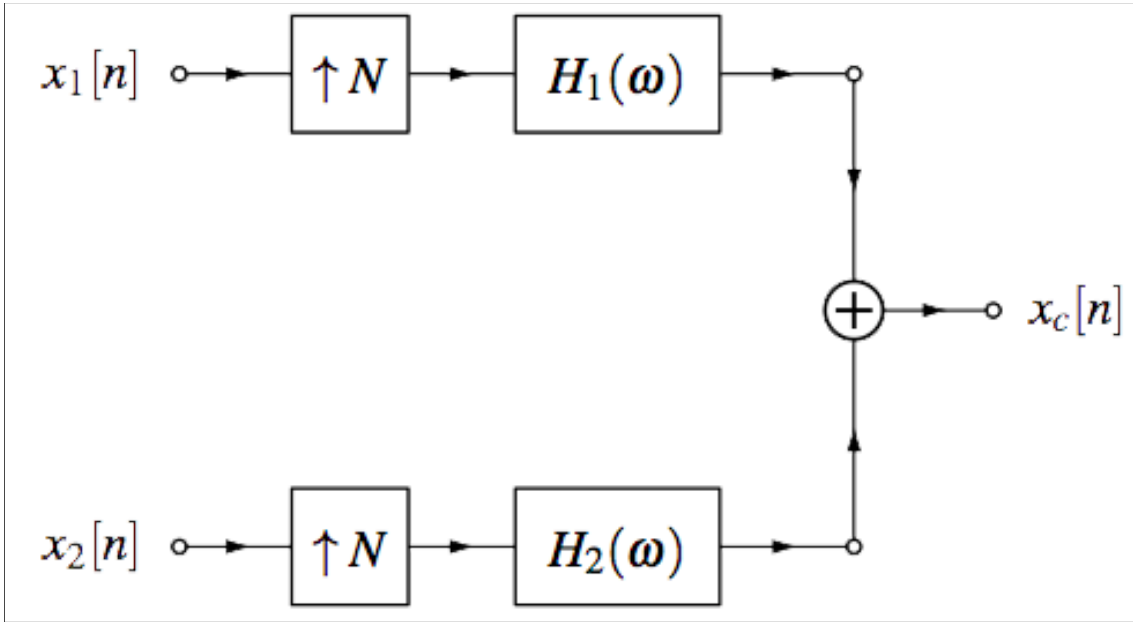


SIMILAR TO TIME-INTERLEAVED ADC's



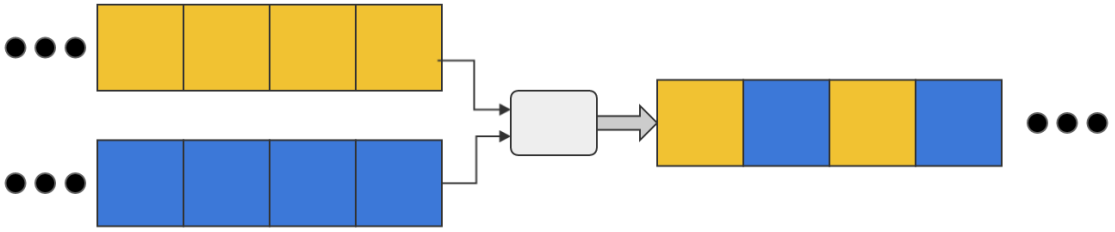
NON-UNIFORM RECONSTRUCTION REQUIRES KNOWING PRECISE TIME-SKEWS

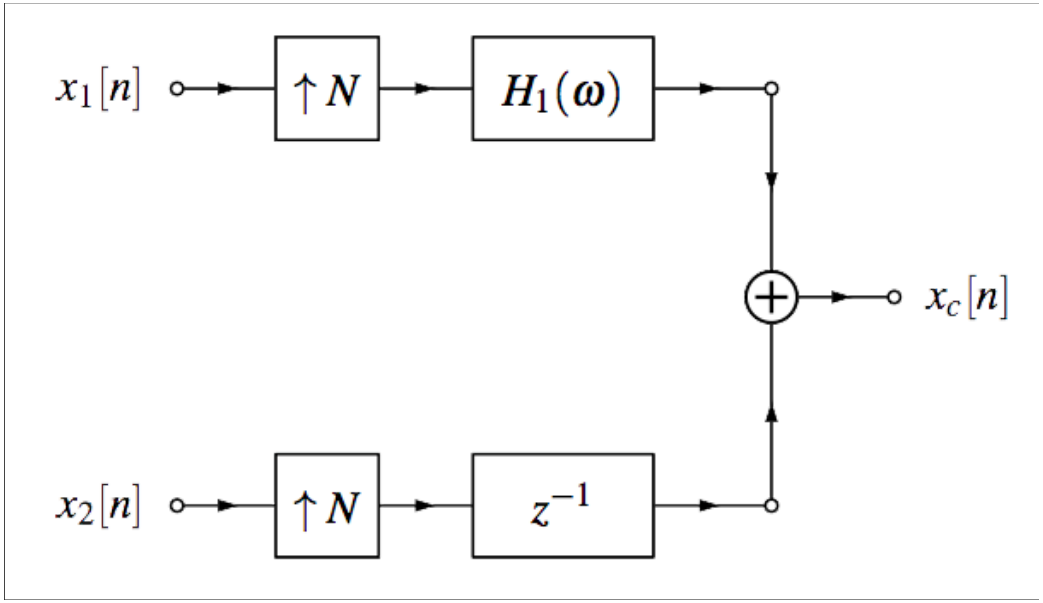
Filterbank interpolation based on Eldar and Oppenheim's paper



PRACTICAL COMPROMISE

Interleaving samples from multiple devices





EVALUATION

SVM	GMM	DTW
15%	5%	65%

Single device

SVM	GMM	DTW
18%	14%	77%

Two devices

- Exhibits improvement over using a single device
- Using even more devices might yield even better results
- Not a proper non-uniform reconstruction

FURTHER ATTACKS

SOURCE SEPARATION

- Use the 3 axes of the gyro
- Learn the number of sound sources around
- Use angle of arrival information for source separation

AMBIENT SOUND RECOGNITION

IS THE USER IN A ROOM/OUTDOORS/ON A STREET?


DEFENSES

SOFTWARE DEFENSES

- Low-pass filter the raw samples
- 0-20 Hz range should be enough for browser based applications (according to WebKit)
- Access to high sampling rate should require a special permission

HARDWARE DEFENSES

- Hardware filtering of sensor signals

  (Not subject to configuration)
- Acoustic masking

  (won't help against vibration of the surface)

CONCLUSION

- Giving applications direct access to hardware is dangerous.

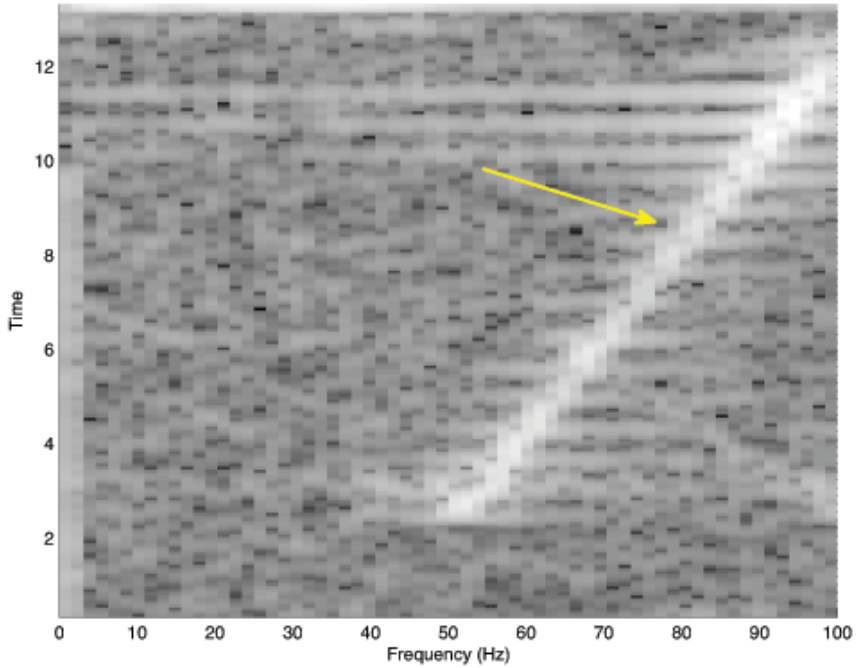
  Especially given the high sampling rate.

THANK YOU VERY MUCH

QUESTIONS?

CRYPTO.STANFORD.EDU/GYROPHONE

IT IS POSSIBLE TO SAMPLE THROUGH JAVASCRIPT



FAQ

- **Did you experiment with an anechoic chamber?**

  Yes, and did not find it beneficial at this stage.

FAQ

- **Perhaps the gyro actually measures the vibrations of the surface?**

  Maybe, but tests suggest it's not only that. In any case it is still dangerous.

FAQ

- **Is it possible to use measurements from multiple devices in other ways?**

  Yes. For example as in MIMO: EGC (Equal Gain Combining).