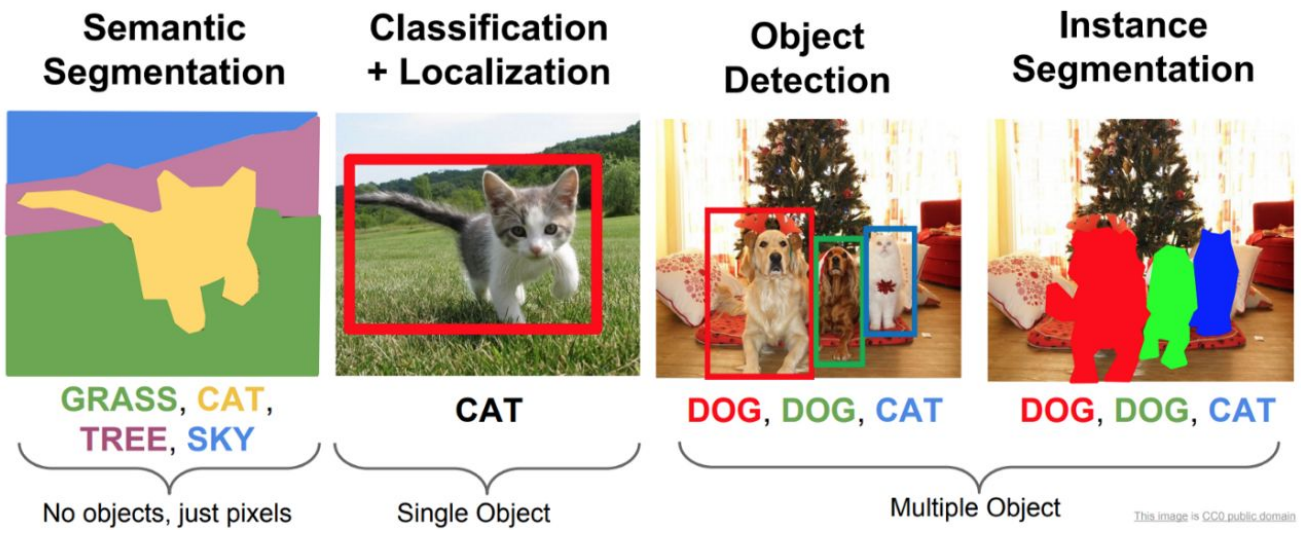
Workshop on
TinyML for
Sustainable Development

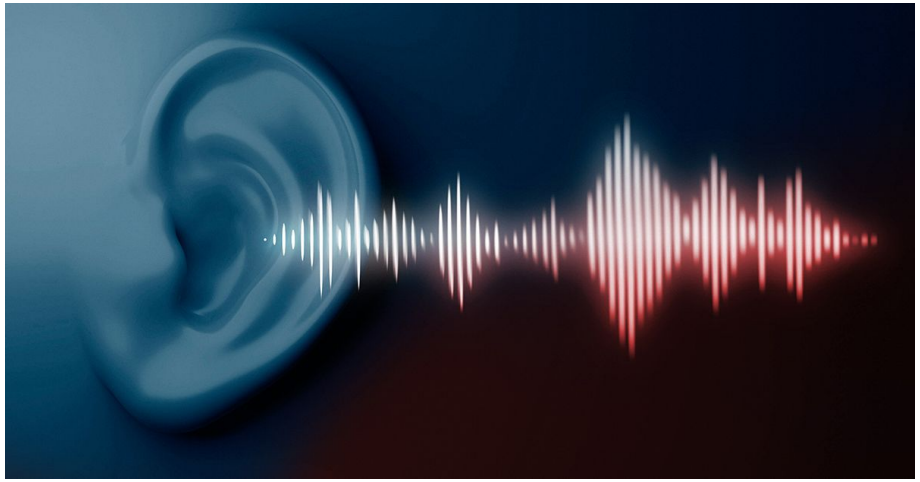
Sound and KWS in the Edge

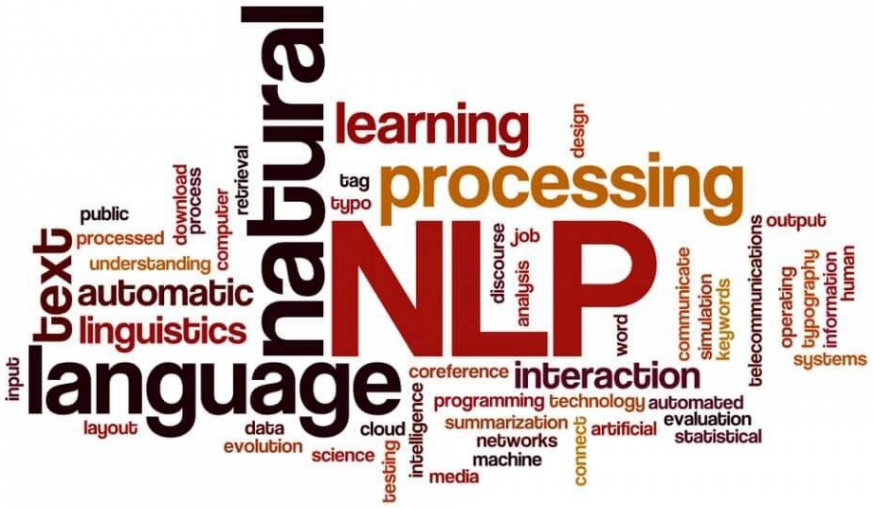
Prof. Jesús Alfonso López
jalopez@uao.edu.co
Universidad Autónoma de Occidente

Images are Ok but, What About Sound?

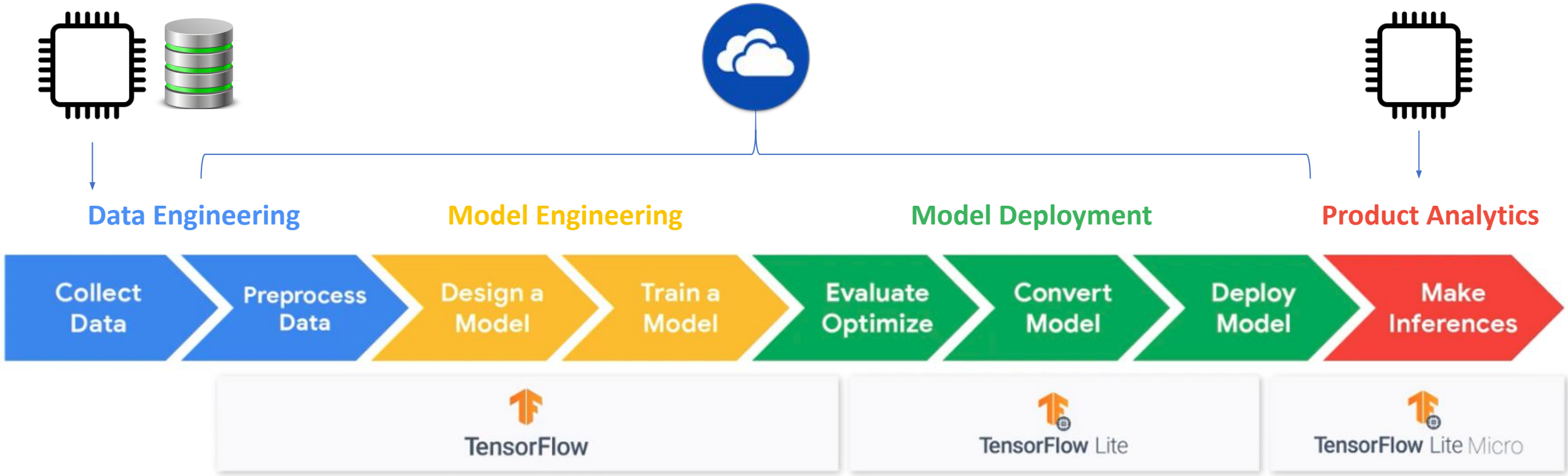
Unstructured Data

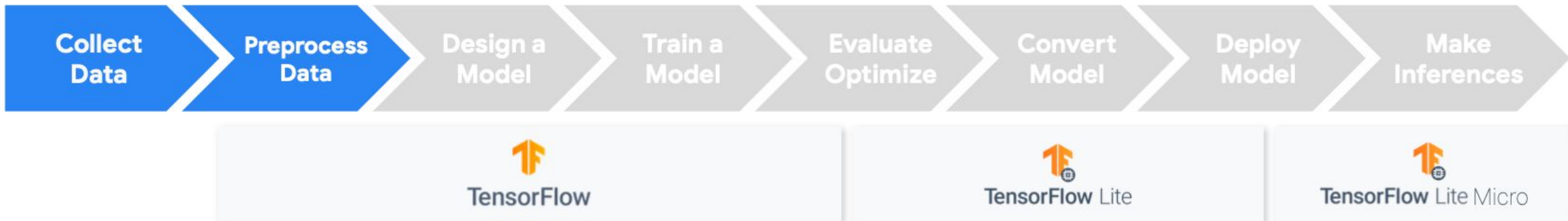




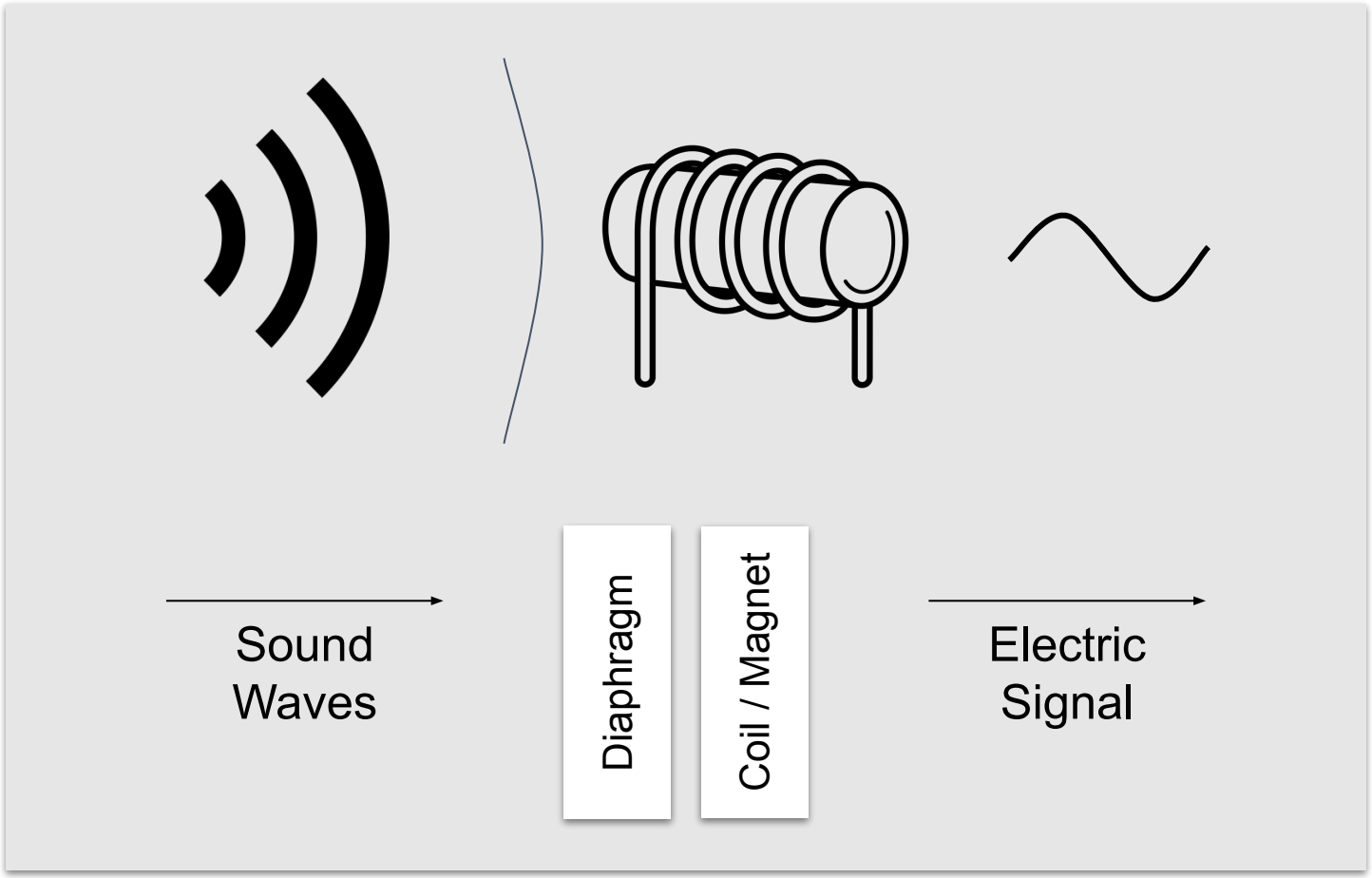


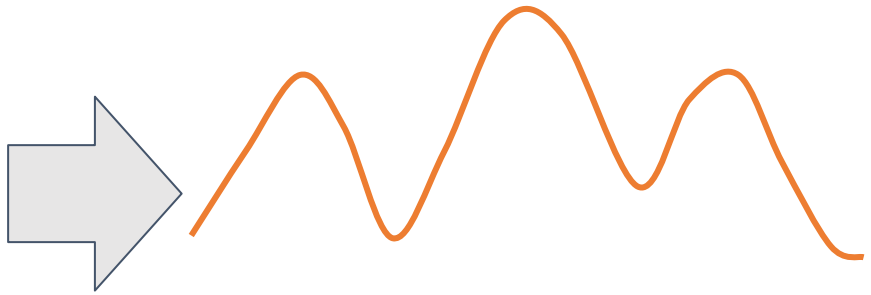
Data Collection & Pre-Processing


Data Engineering
Model Engineering
Model Deployment
Product Analytics
Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro

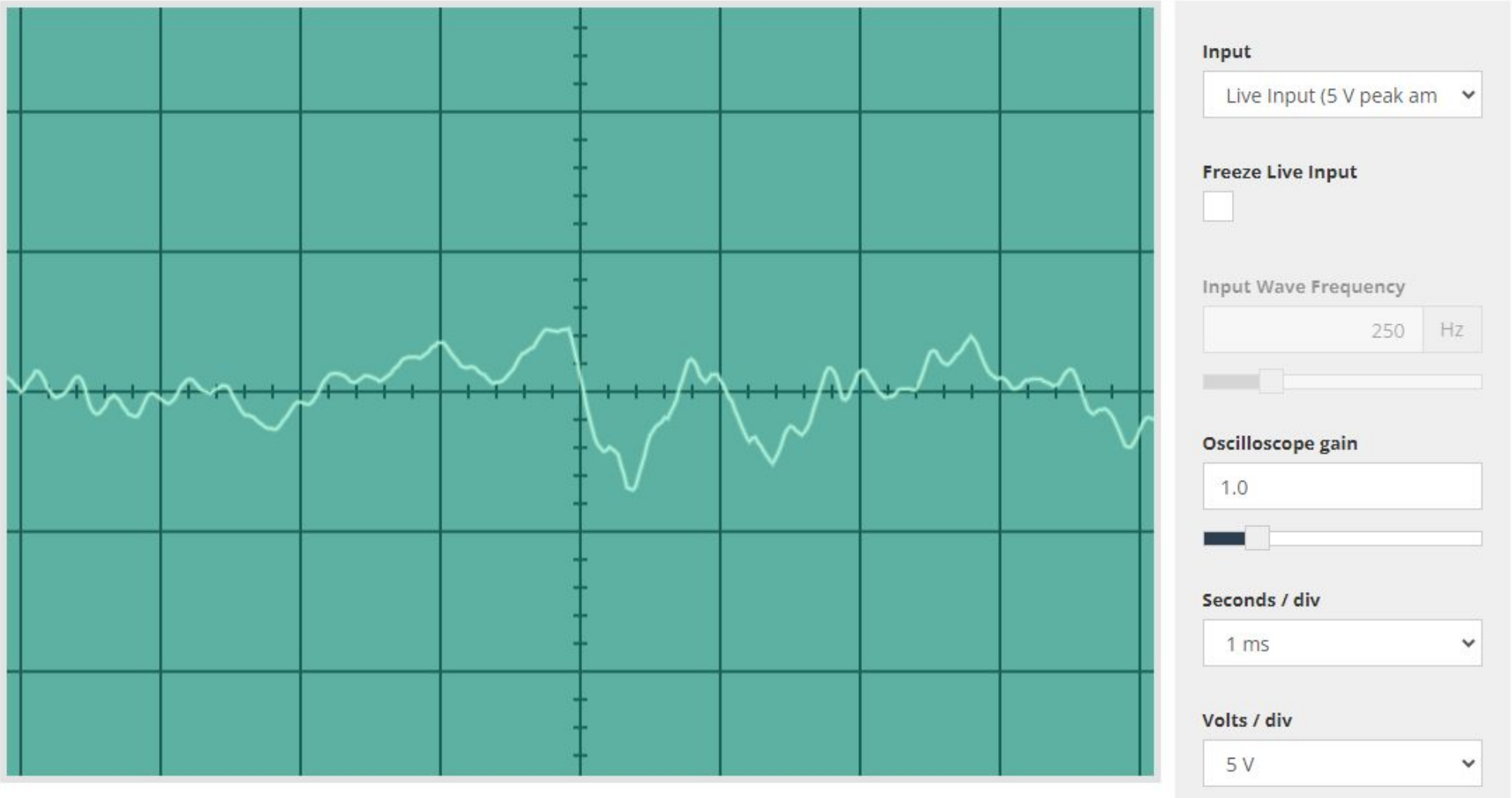
Sensor Data





16kHz signal

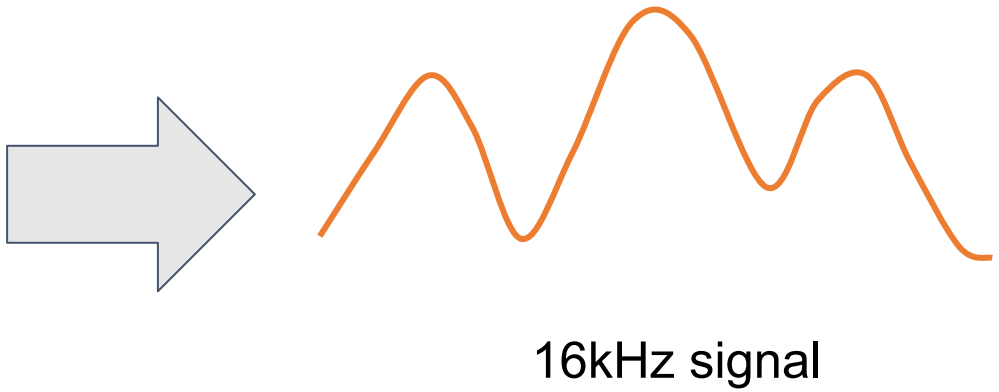
Sensor Data



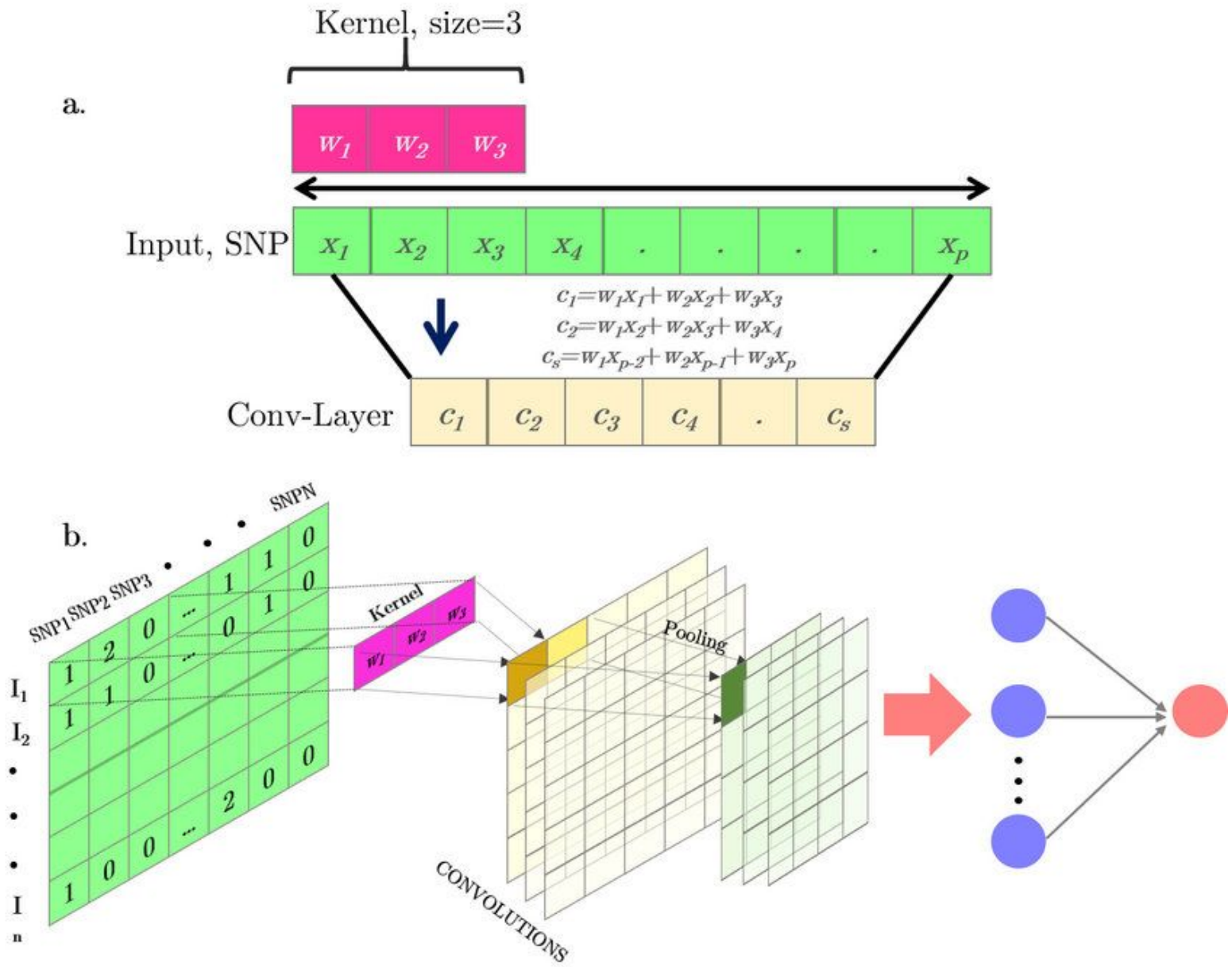
https://academo.org/demos/virtual-oscilloscope/

Sensor Data

- 1D signal
- 16kHz signal, so that's **16000** samples (points / second)
- How do you feed ***all*** of that data into the model?
- Need to **think creatively** about the input signal!



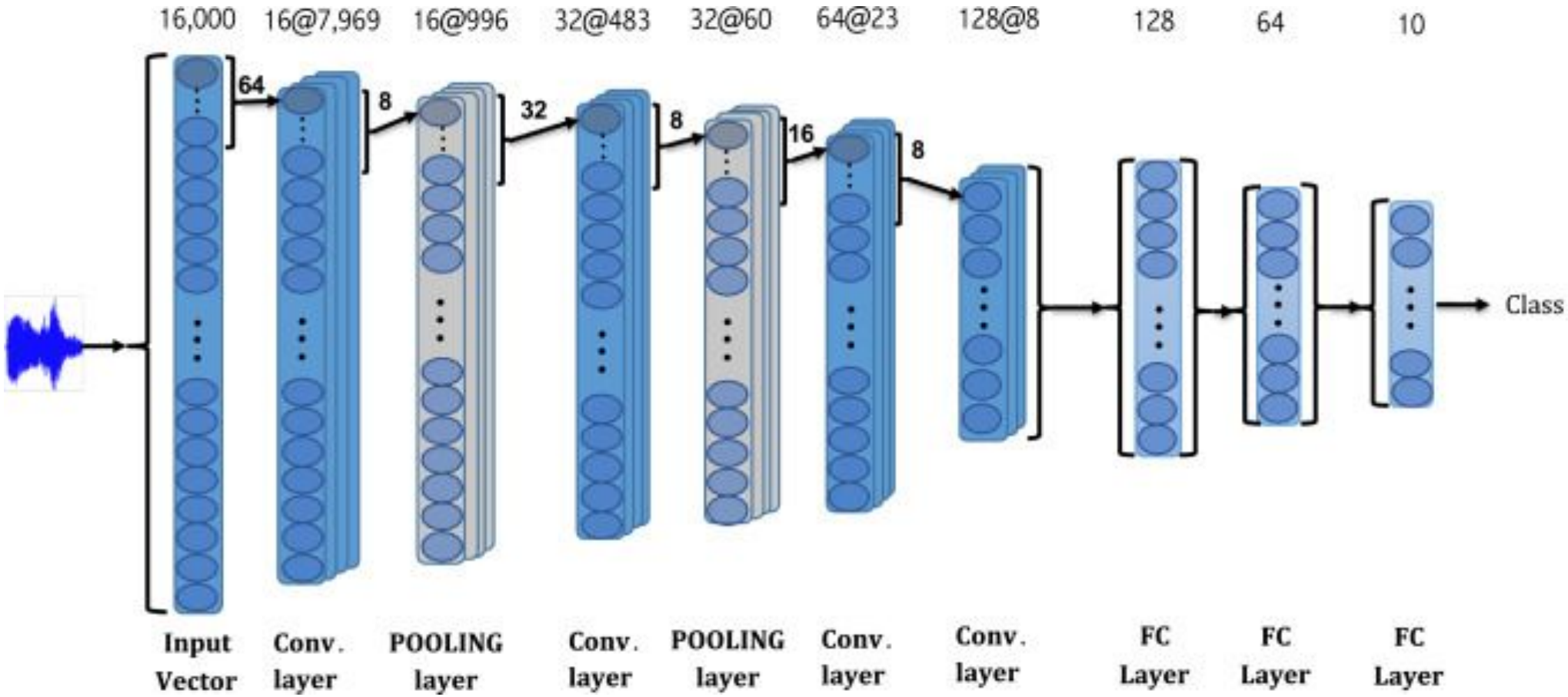
1D Convolution



https://www.researchgate.net/figure/a-Simple-scheme-of-a-one-dimension-1D-convolutional-operation-b-Full_fig2_334609713

1D Convolution

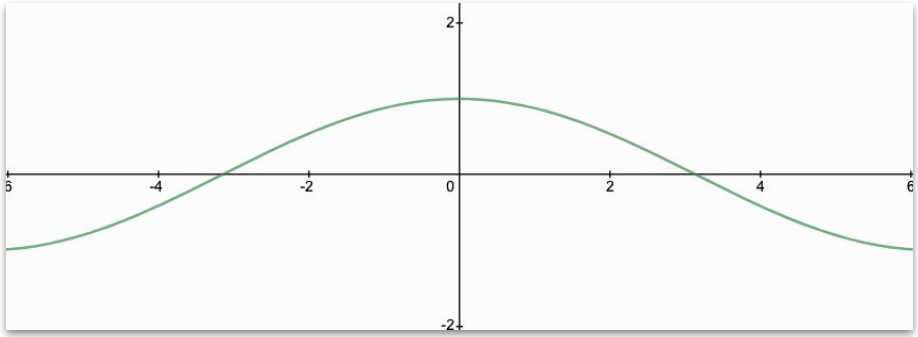
Raw Data



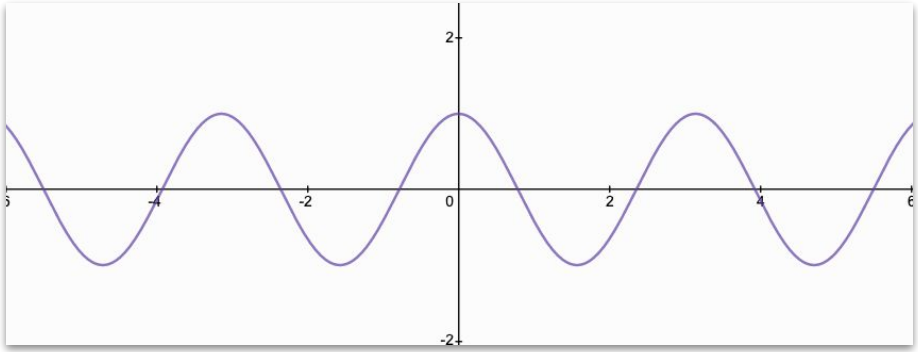
End-to-End Environmental Sound Classification using a 1D Convolutional Neural Network

https://arxiv.org/pdf/1904.08990

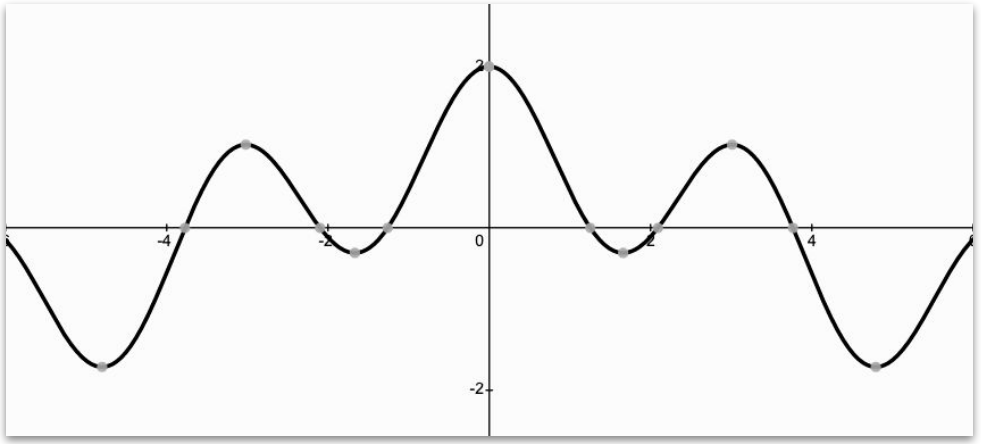
Signal **Components**?



+



=



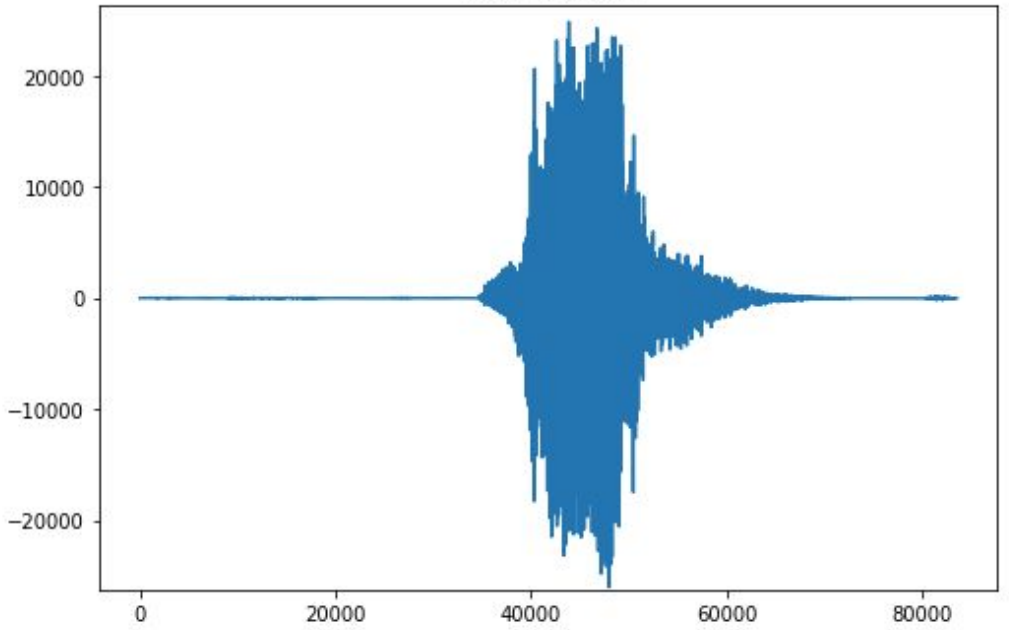
Signal **Components**?

?

+

?

=



Signal **Components**?

R// Fourier Transform

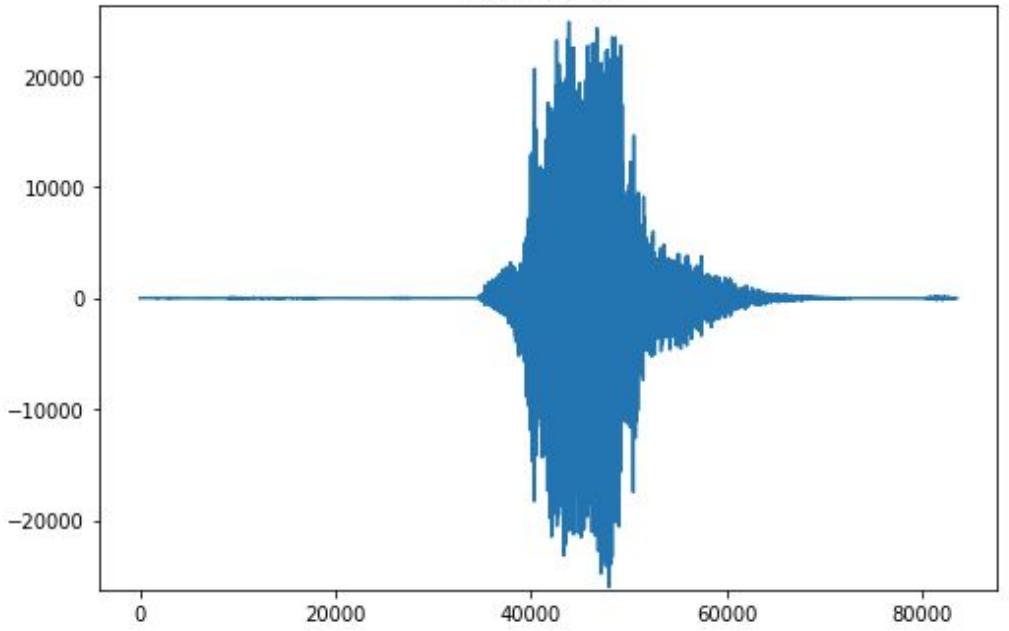
https://www.youtube.com/watch?v=spUNpyF58BY&sttick=0

?

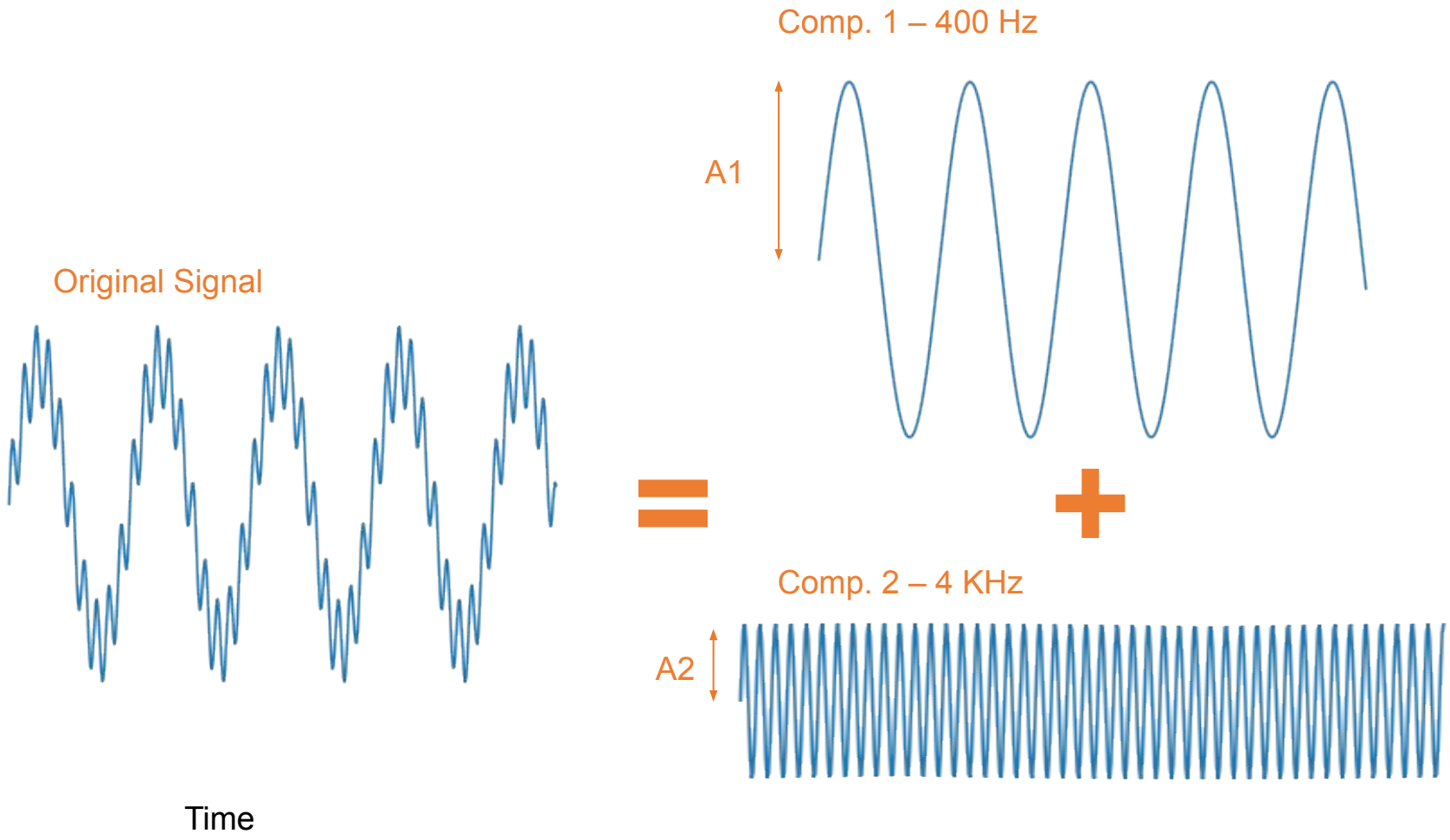
+

?

=



Fourier Transform

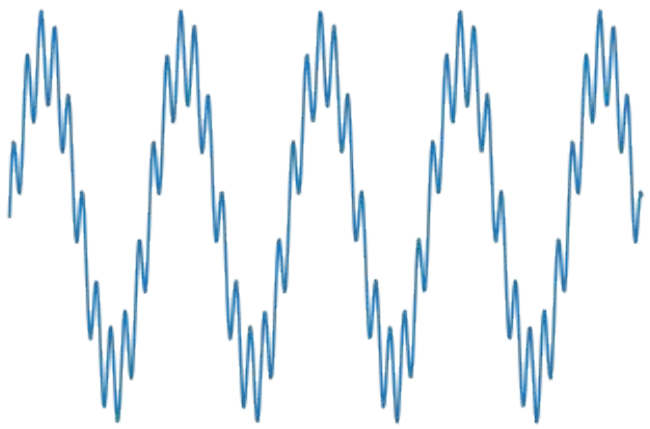


https://www.youtube.com/watch?v=spUNpyF58BY&sttick=0

Fourier Transform

```
from scipy.fft import fft
yf = fft(raw signal)
plt.plot(xf, np.abs(yf));
```

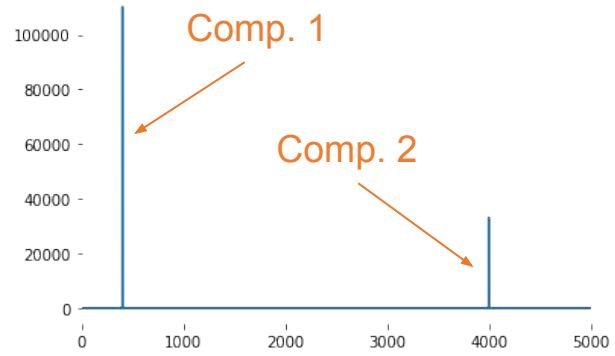
Original Signal



Time

$$F(j\omega) = \int_{-\infty}^{\infty} f(t)e^{-j\omega t}dt$$

Fourier Transform

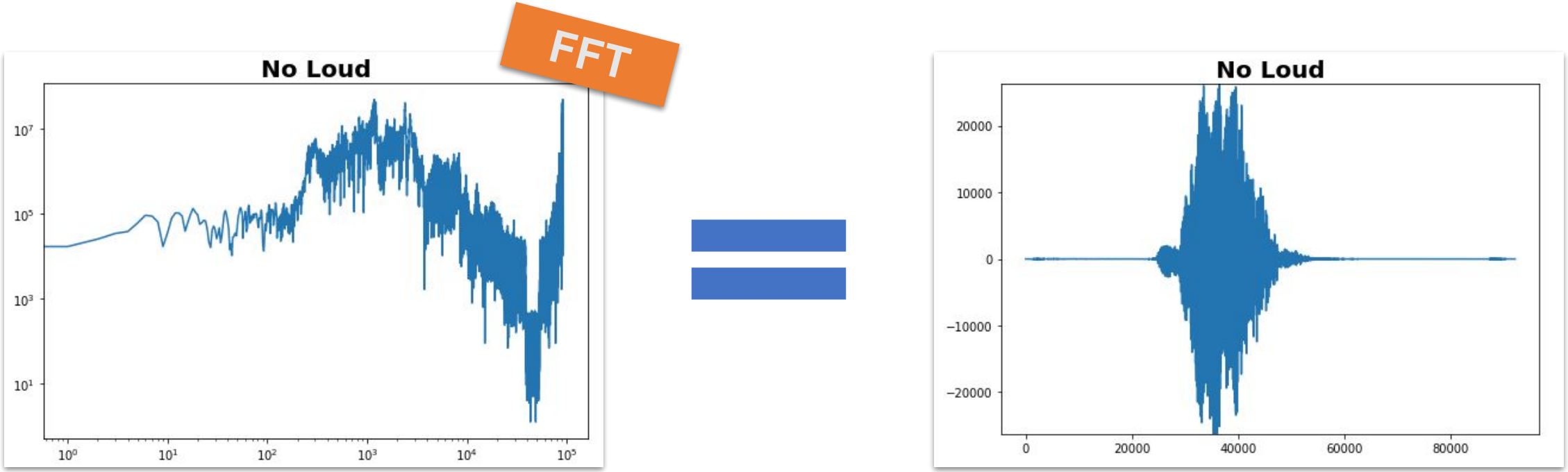


Frequency

https://prajwalsouza.github.io/Experiments/Fourier-Transform-Visualization.html

https://www.youtube.com/watch?v=spUNpyF58BY&sttick=0

Signal **Components**!

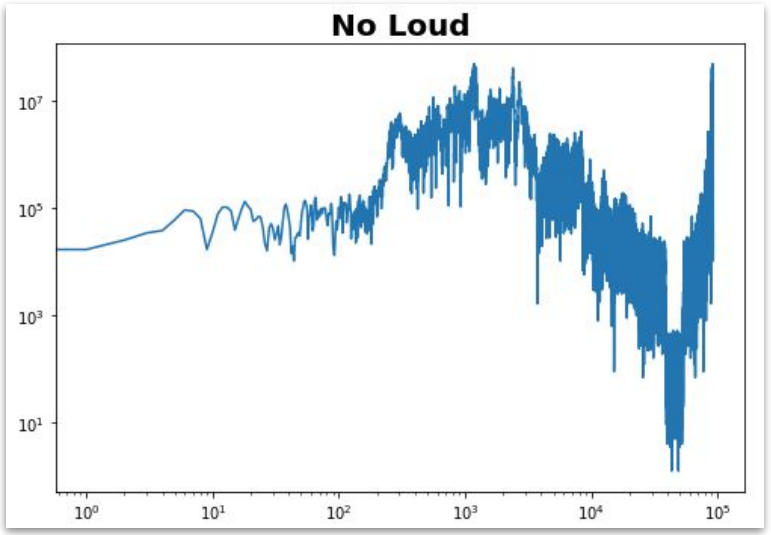


Frequency

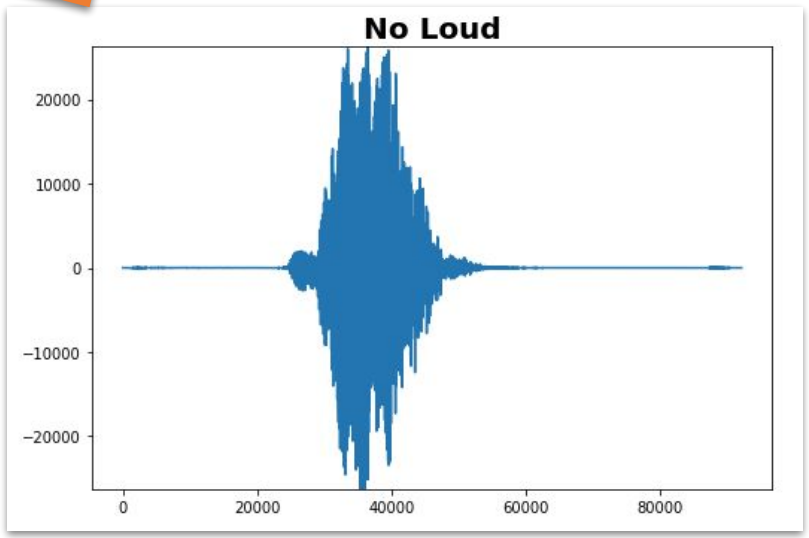
Time

How Can We Mix Time and Frequency Information?

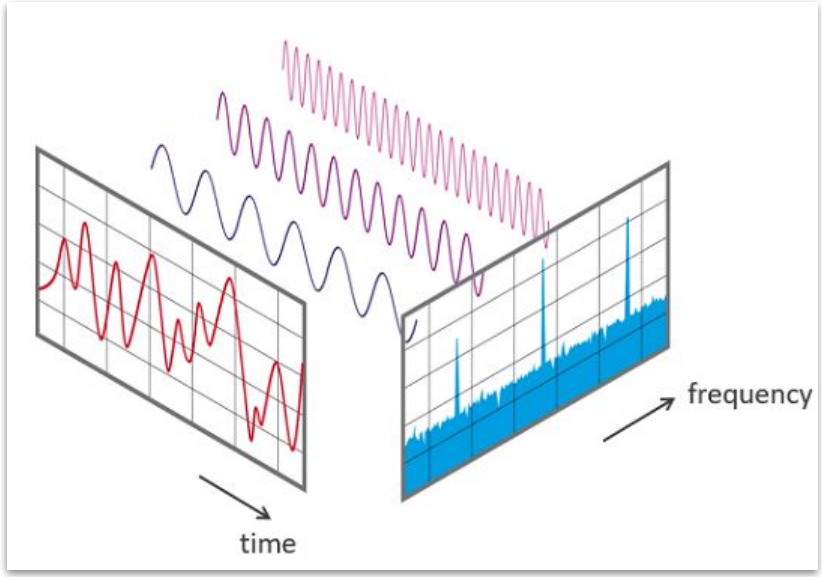
FFT



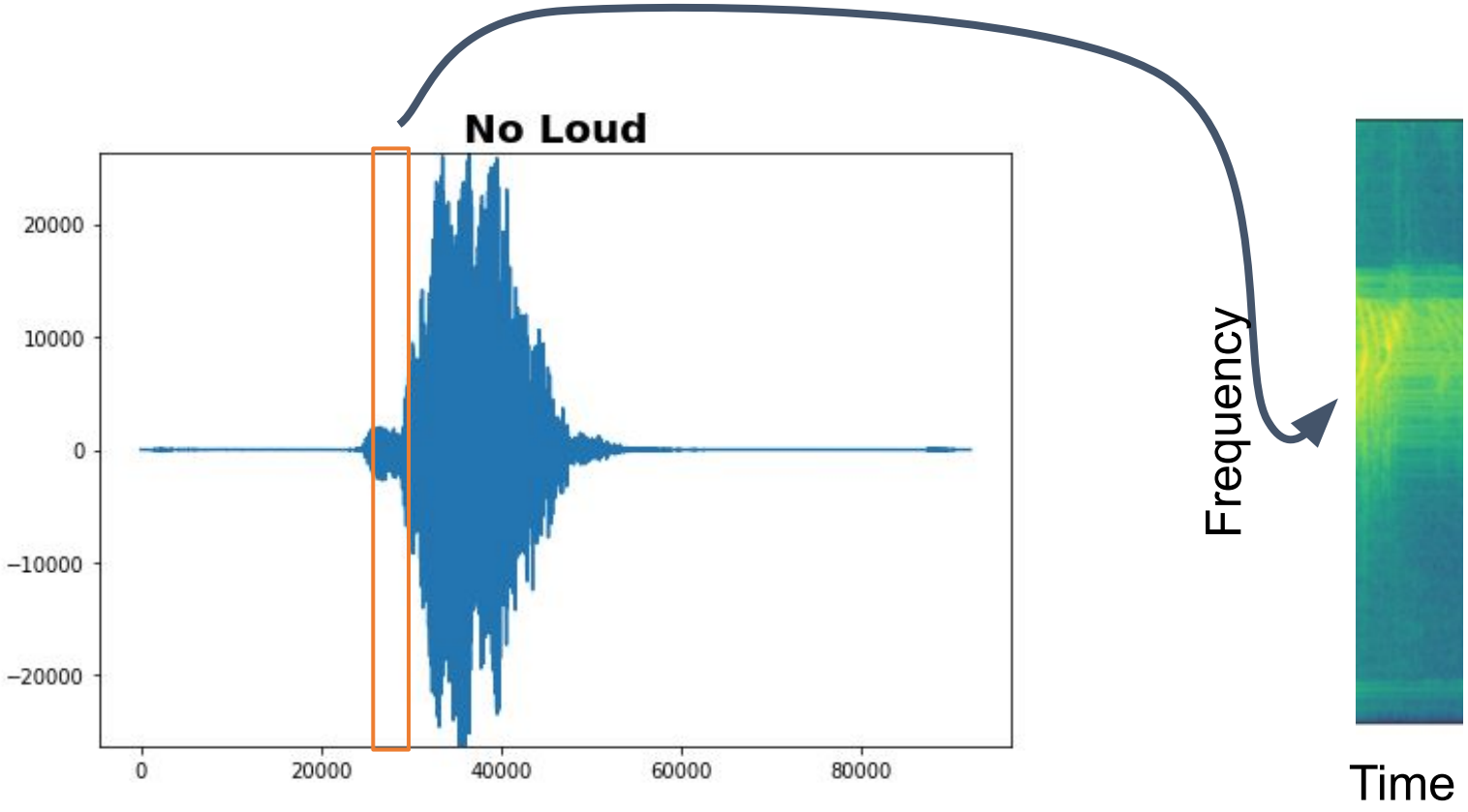
Frequency



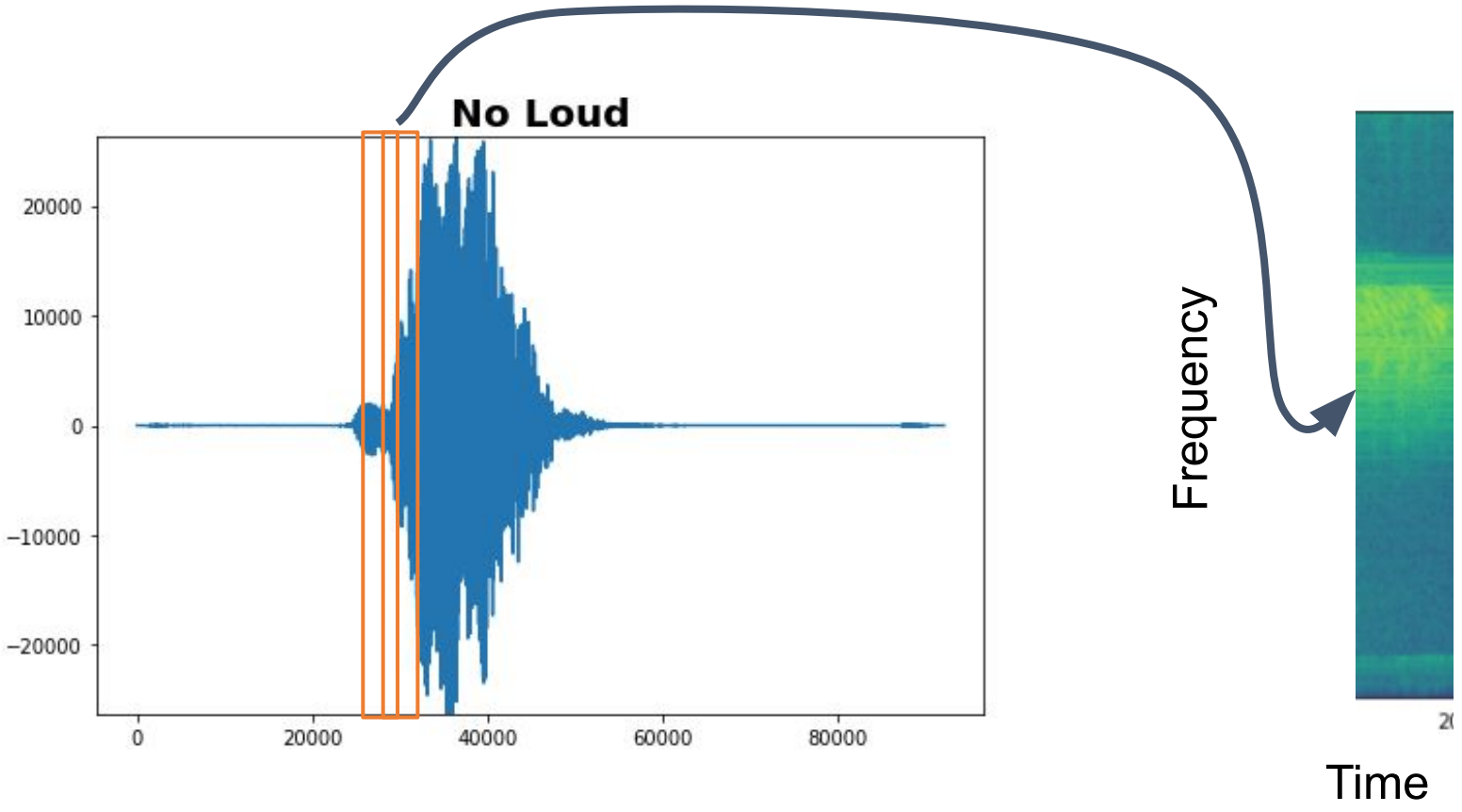
Time



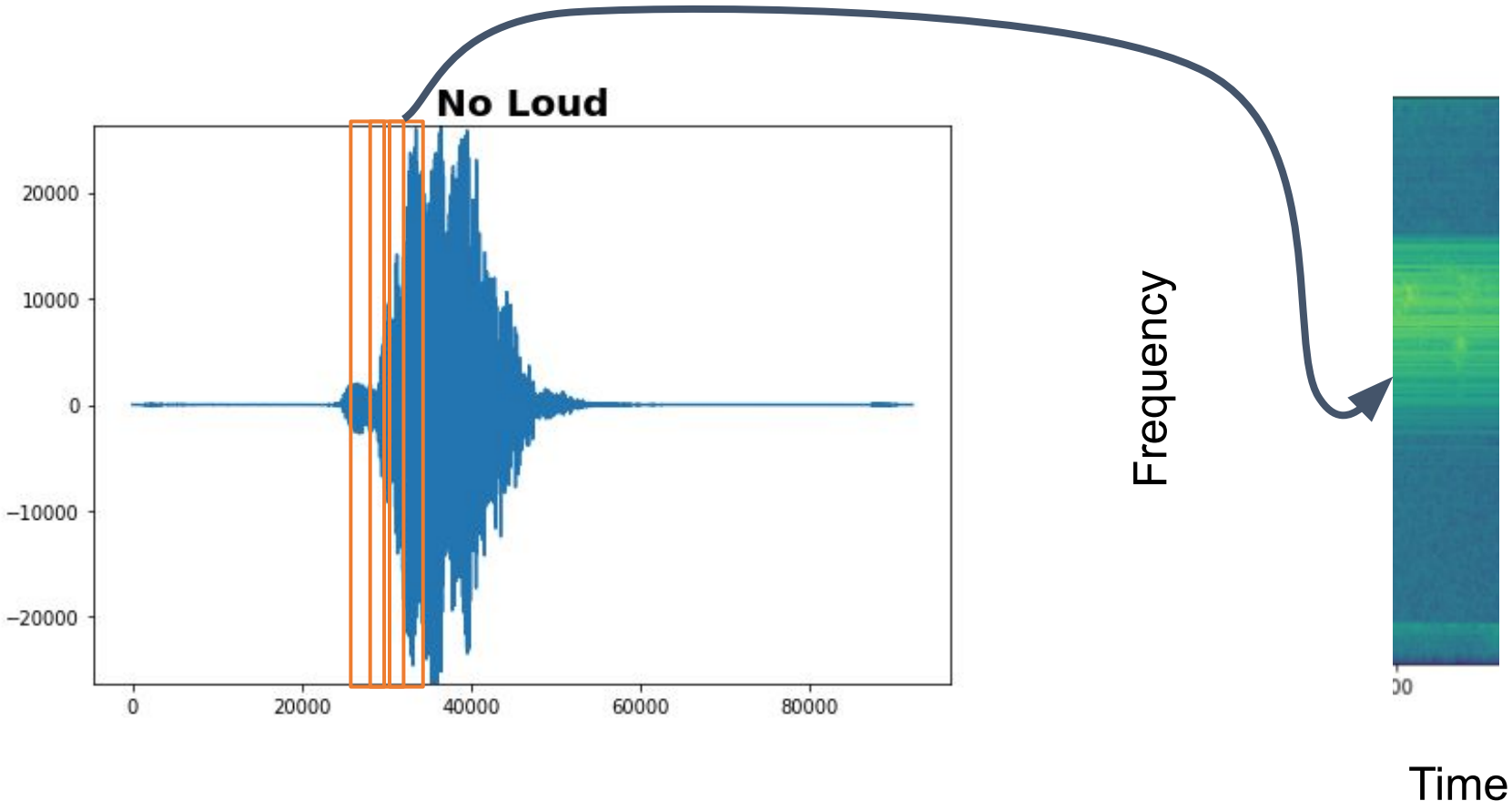
Data Preprocessing



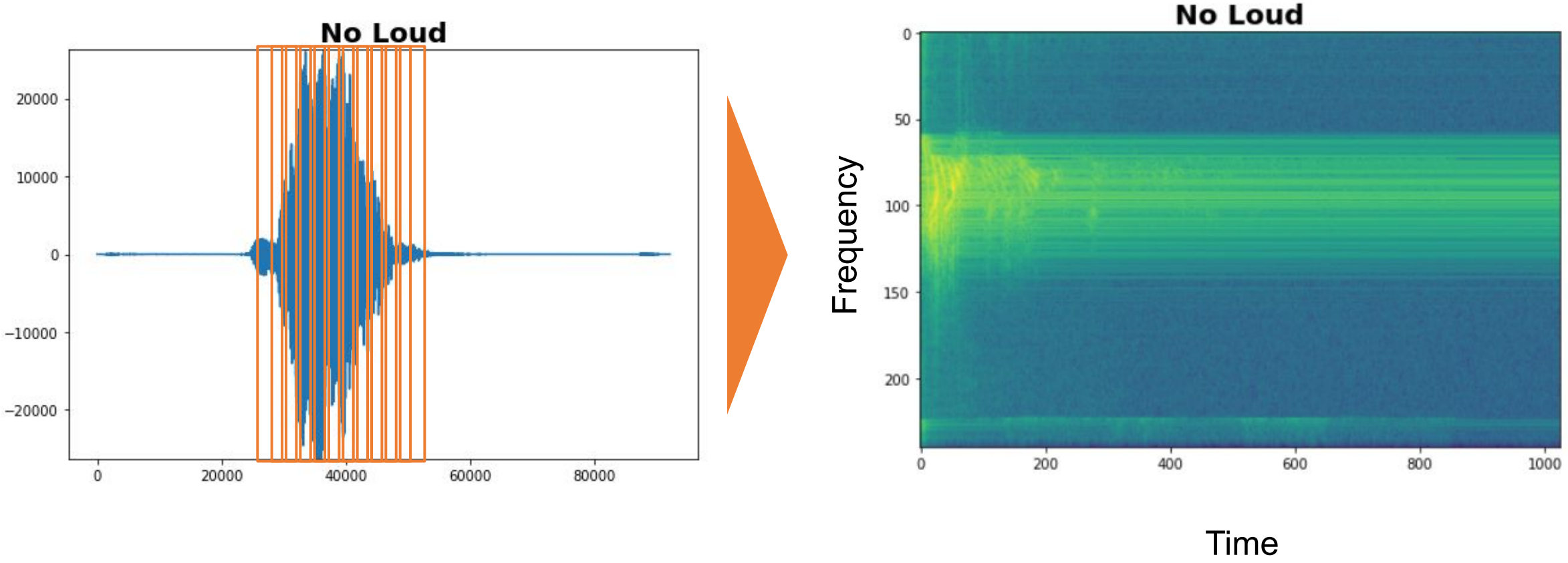
Data Preprocessing



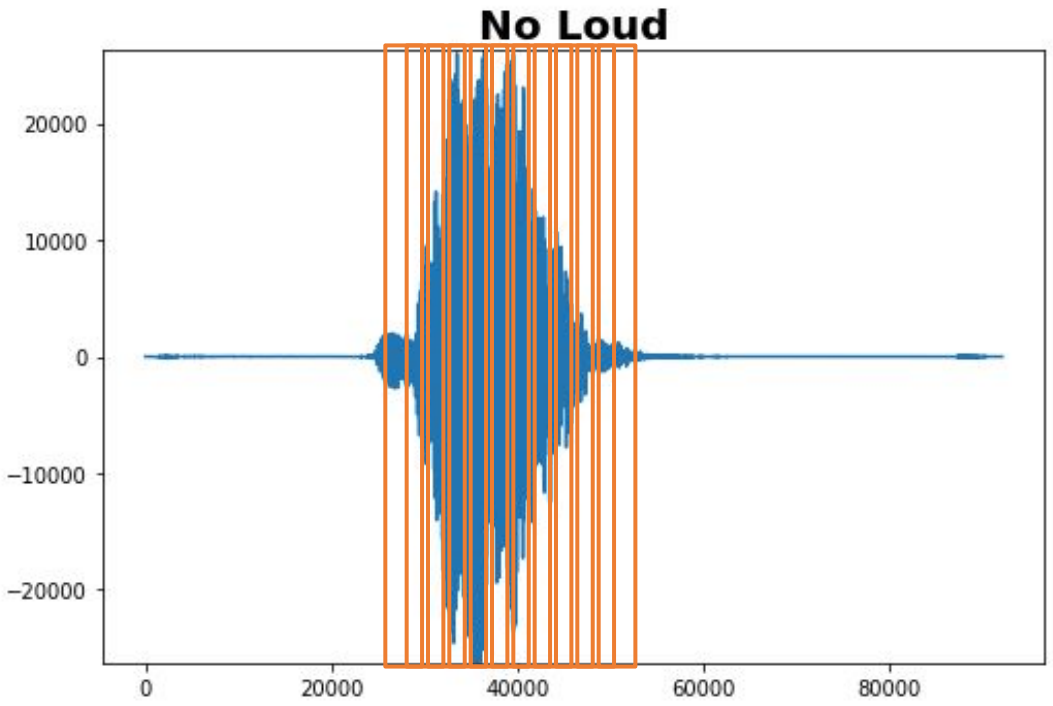
Data Preprocessing

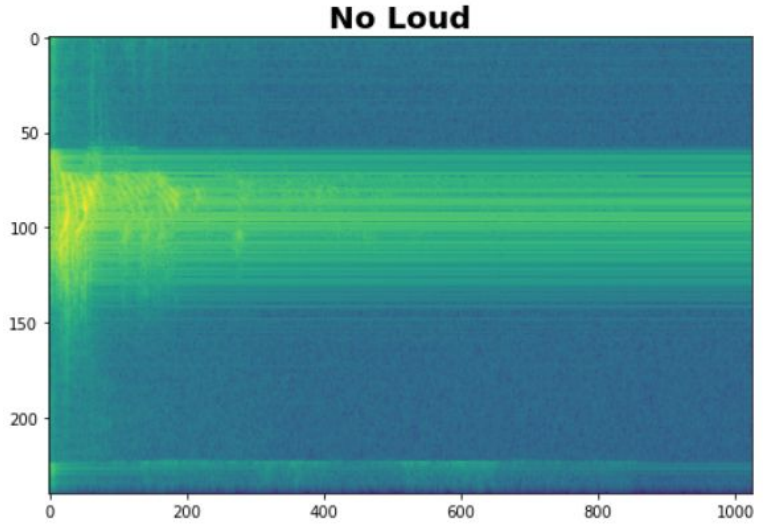


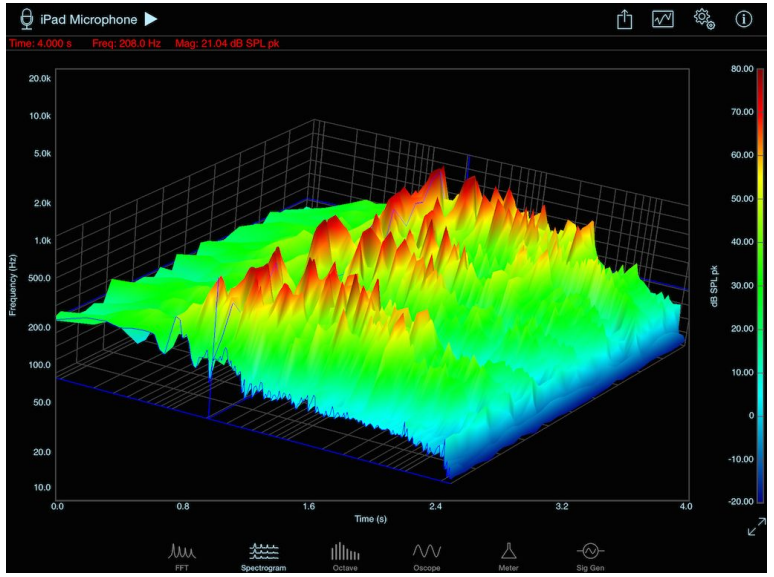
Data Preprocessing: **Spectrograms**



Data Preprocessing: **Spectrograms**

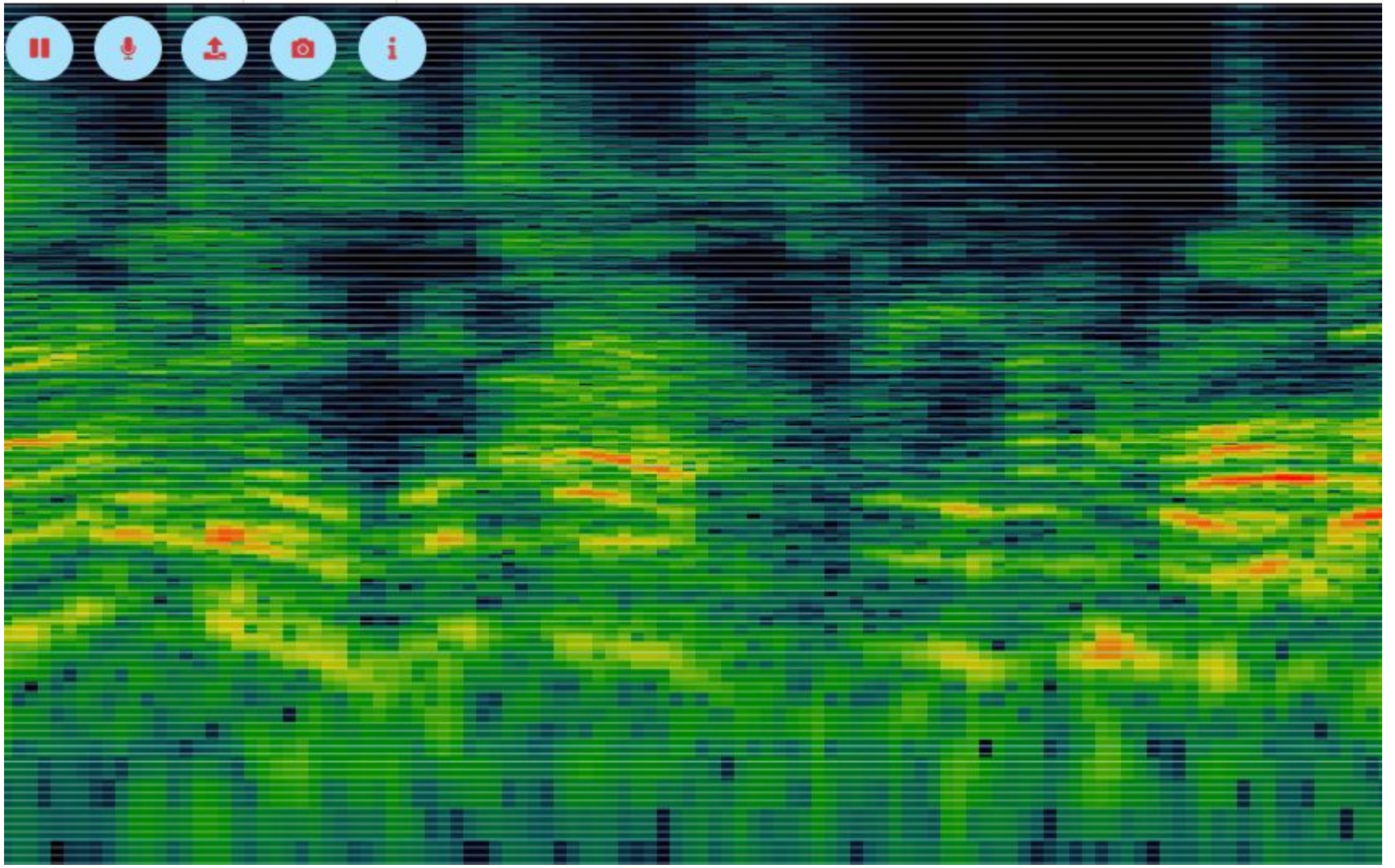




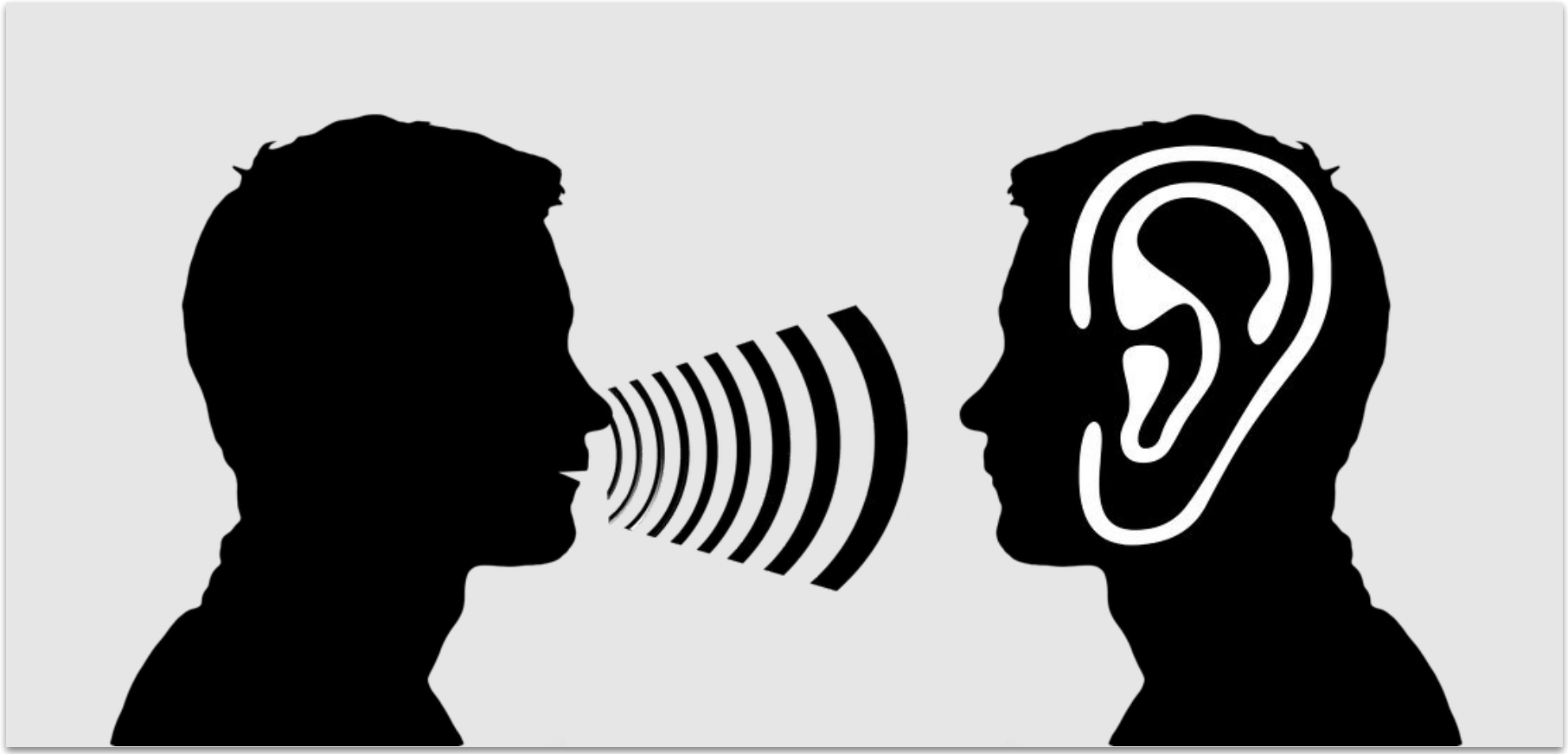


Data Preprocessing: **Spectrograms**

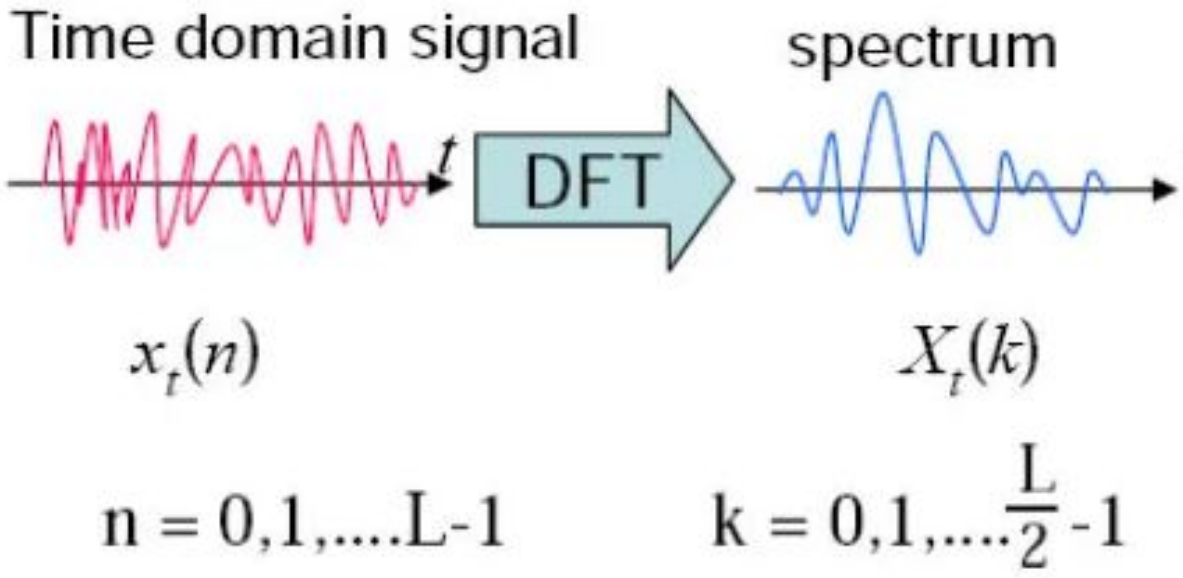
https://spectrogram.sciencemusic.org/

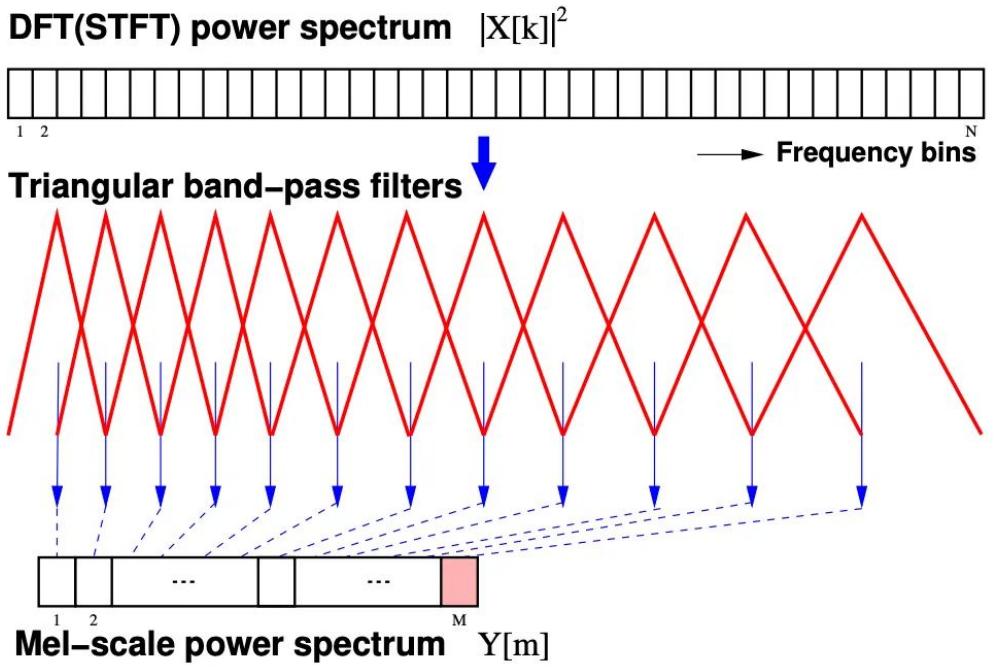


Can we find more
salient features?



Mel-Frequency Cepstral Coefficients : **MFCC**



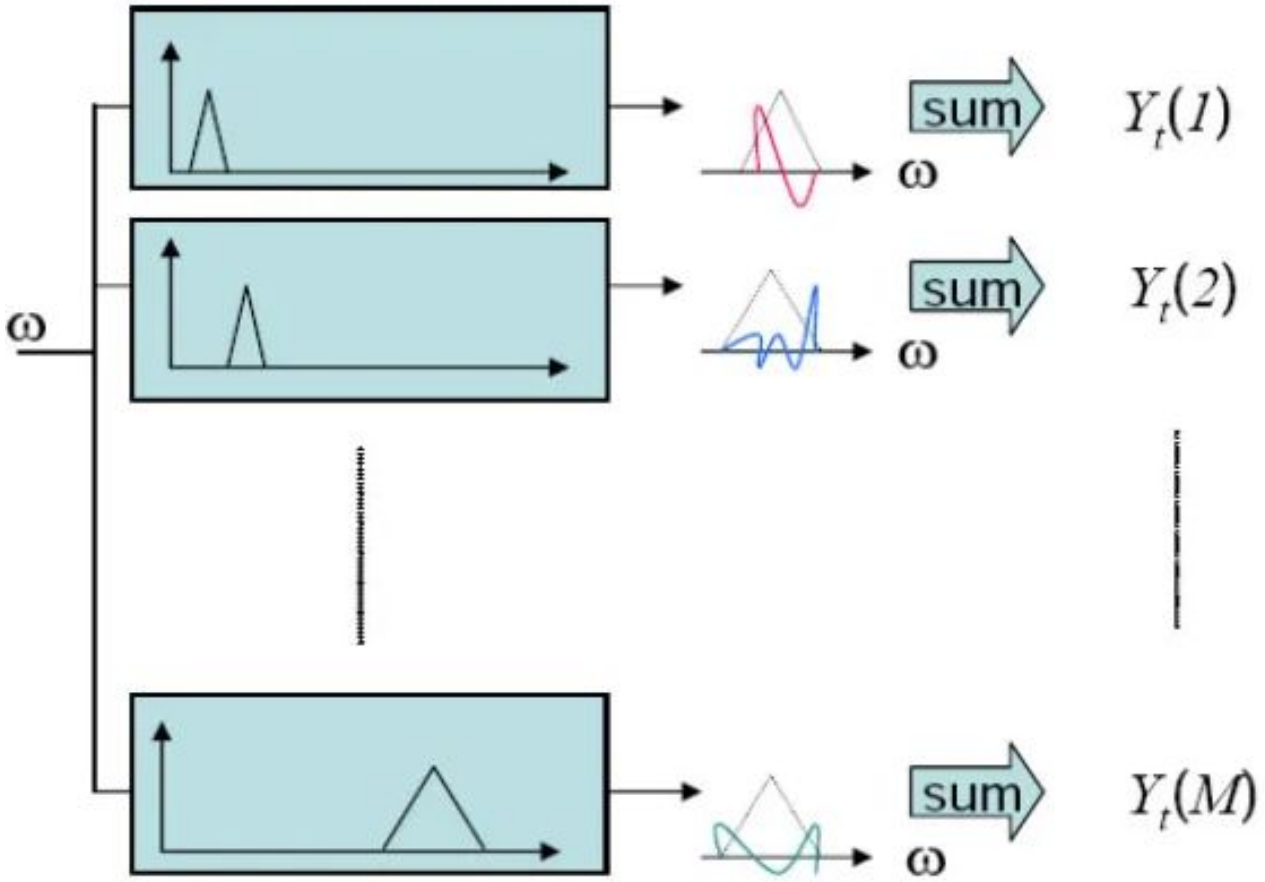


https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9

Mel-Frequency Cepstral Coefficients : **MFCC**

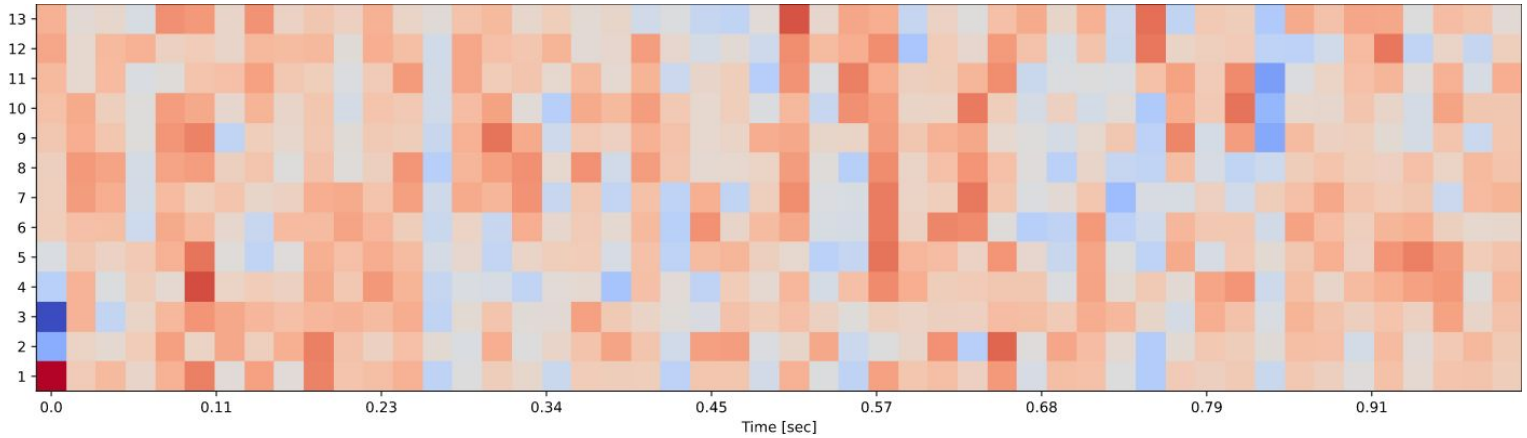
$$Y_t[m] = \sum_{k=1}^{N} W_m[k]\,|X_t[k]|^2$$

where k : DFT bin number $(1, \ldots, N)$
m : mel-filter bank number $(1, \ldots, M)$



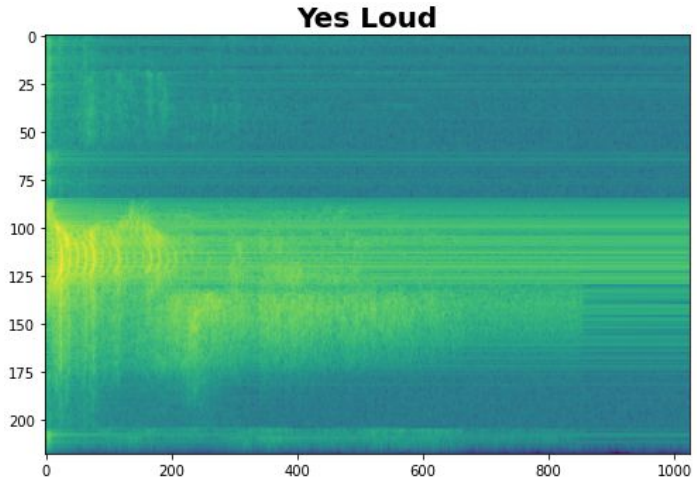
https://jonathan-hui.medium.com/speech-recognition-feature-extraction-mfcc-plp-5455f5a69dd9

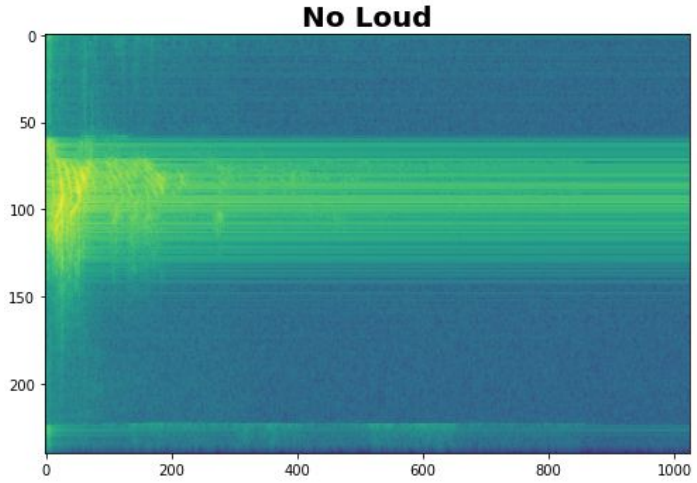
Mel-Frequency Cepstral Coefficients : **MFCC**

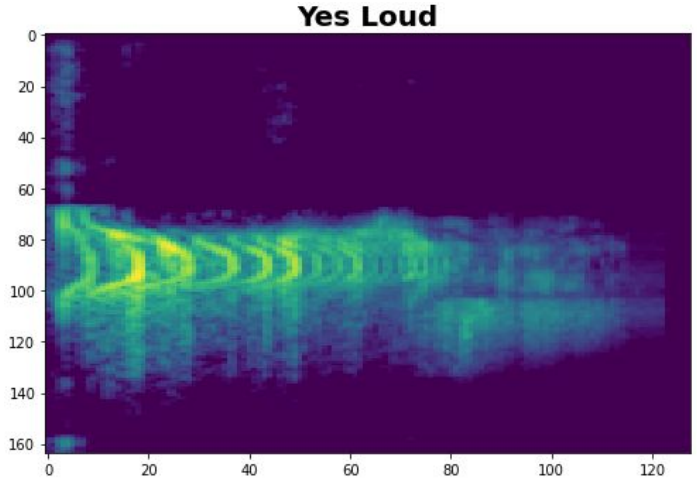


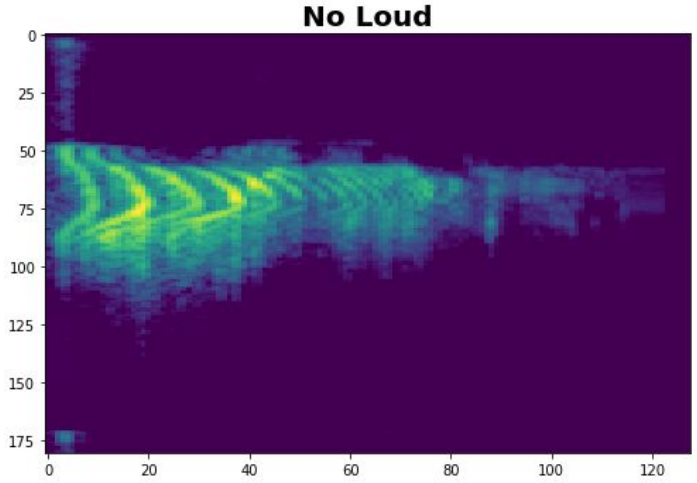
Ejemplo de MFCC

Spectrograms vs MFCCs

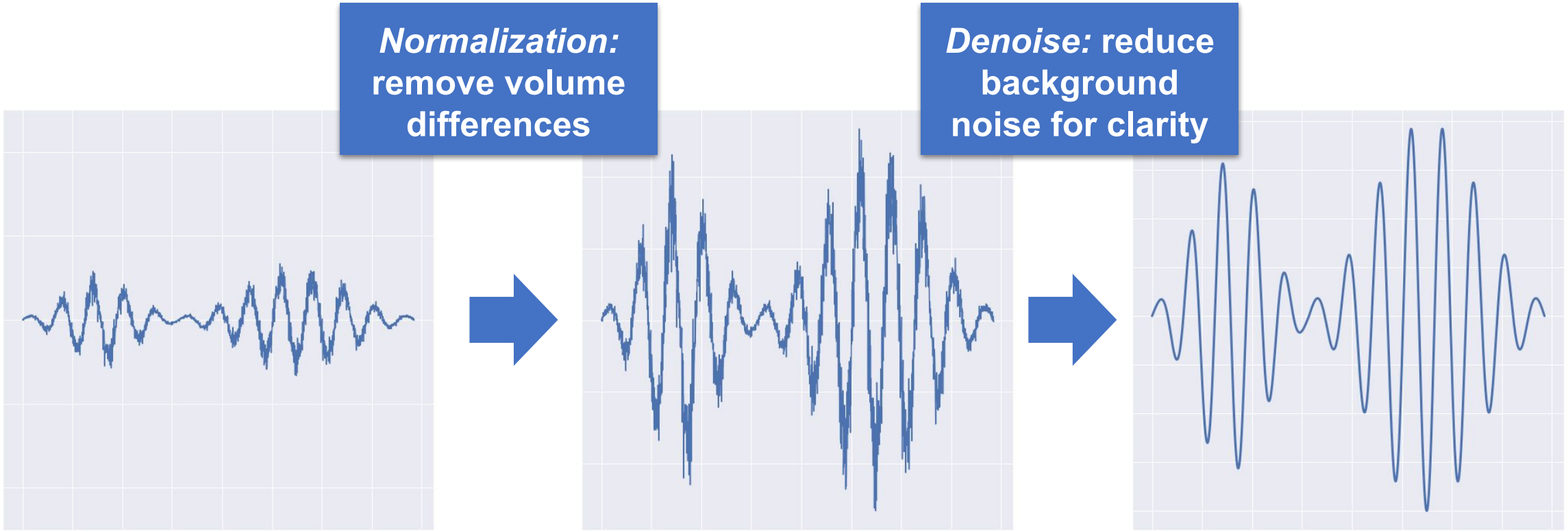








Additional **Feature Engineering**



What is KeyWord Spotting?

Keyword Spotting v. General Speech Recognition

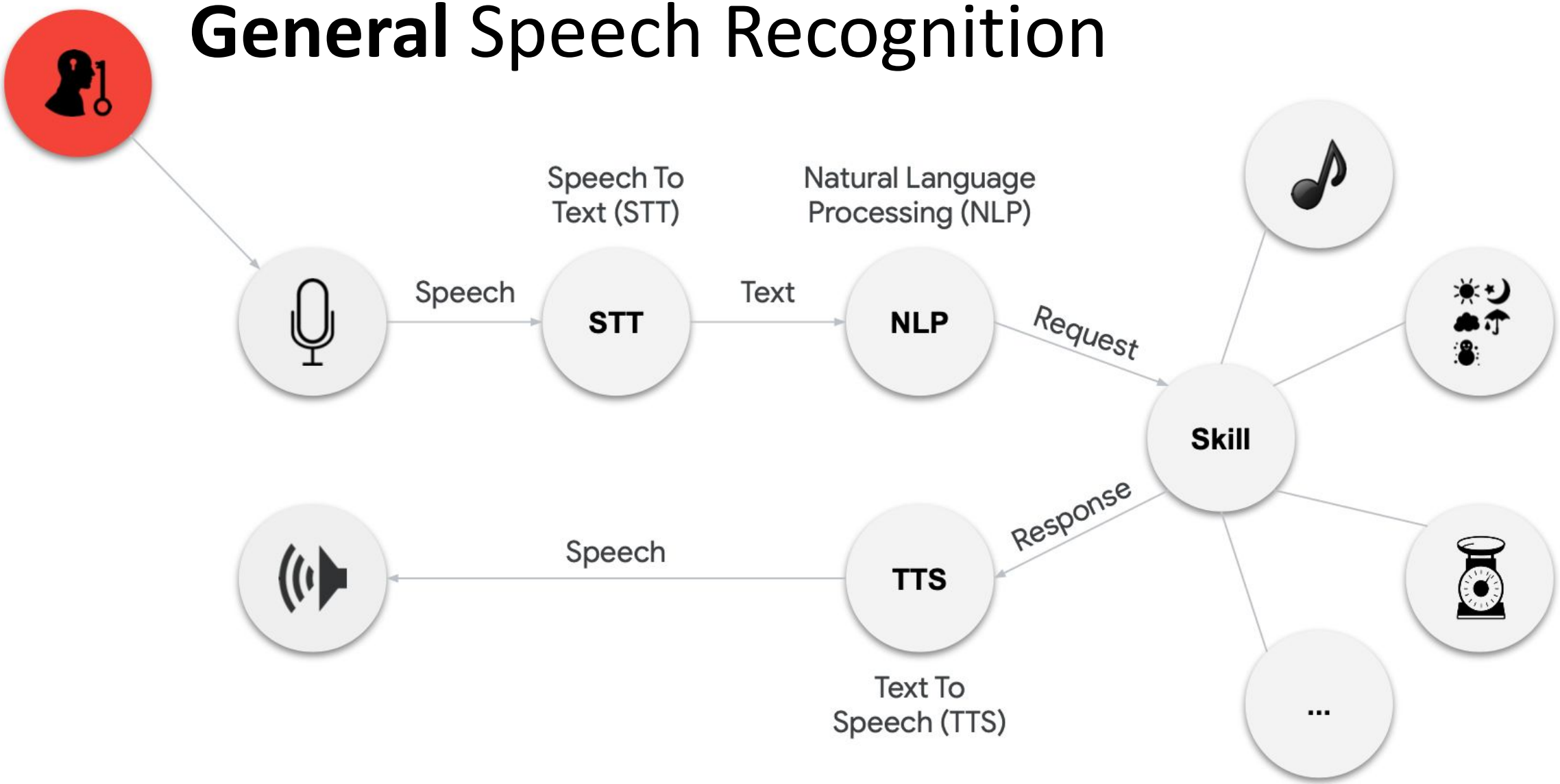
- **Keyword spotting** is one of the most successful examples of **TinyML**
  - Low-power, continuous, on-device
  - Common Voice SWTS* expands keyword spotting to more languages

    * **S**ingle **W**ord **T**arget **S**egment

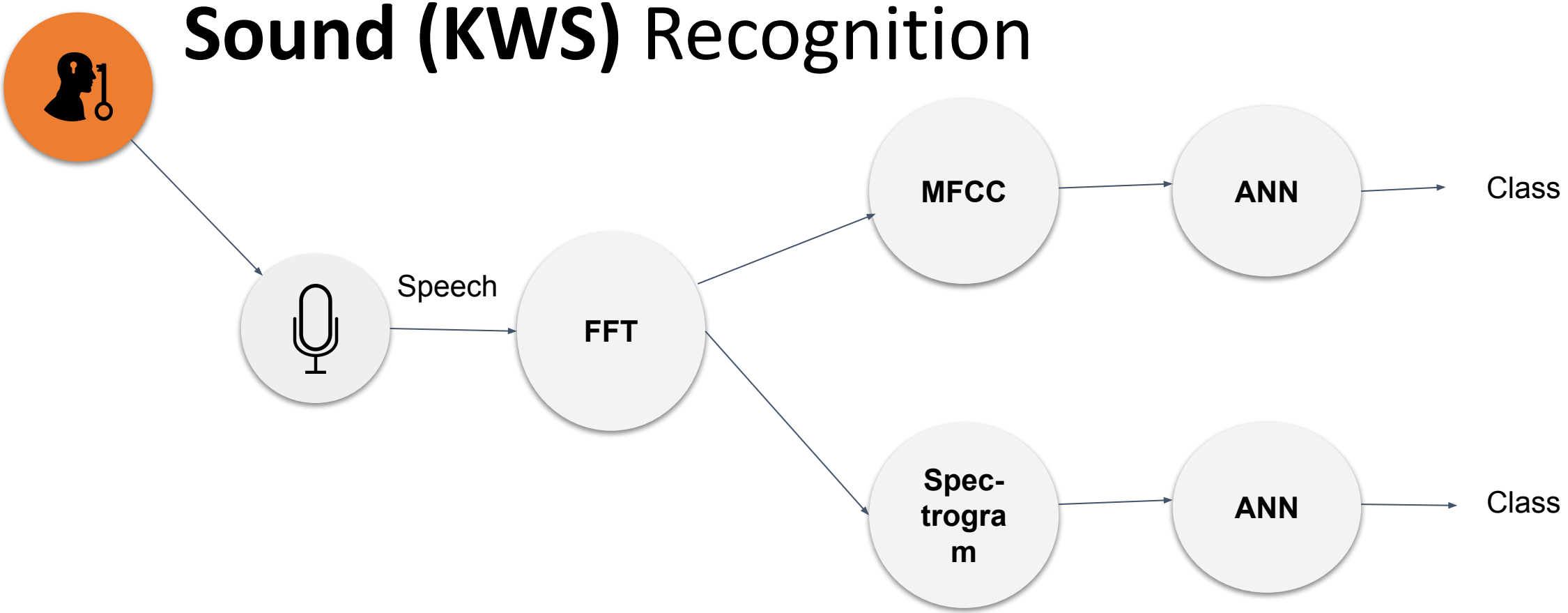
- **General ASR*** still requires **larger, power-hungry models**
  - But it can run on mobile devices (offline dictation on smartphones)

    * Automatic Speech Recognition

General Speech Recognition



Sound (KWS) Recognition




amazon


amazon


amazon

More than just voice

- Secutity (Broken Glass)
- Industry (Anomaly Detection)
- Medical (Snore, Toss)
- Nature (Bee, insect sound)

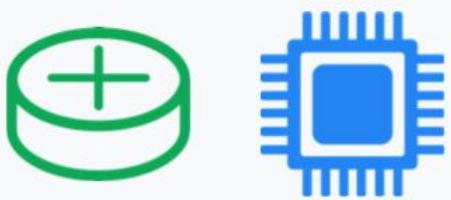
Keyword Spotting
Challenges/Constrains

Challenges and Constraints

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy



Battery & Memory

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

LATENCY

Provide results quickly, respond in real-time to the user

Challenges and Constraints

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

BANDWIDTH

Minimize data sent over the network (slow and expensive)

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

ACCURACY

Listen continuously,
but only trigger at the right time

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

PERSONALIZATION

Trigger for the user and **not** for background noise

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

SECURITY

Safeguarding the data that is being sent to the cloud

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

PRIVACY

Safeguarding the data that is being sent to the cloud

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

Security & Privacy

Battery & Memory

BATTERY

Limited energy, operate on coin-cell type batteries

Challenges and **Constraints**

Latency & Bandwidth

Accuracy & Personalization

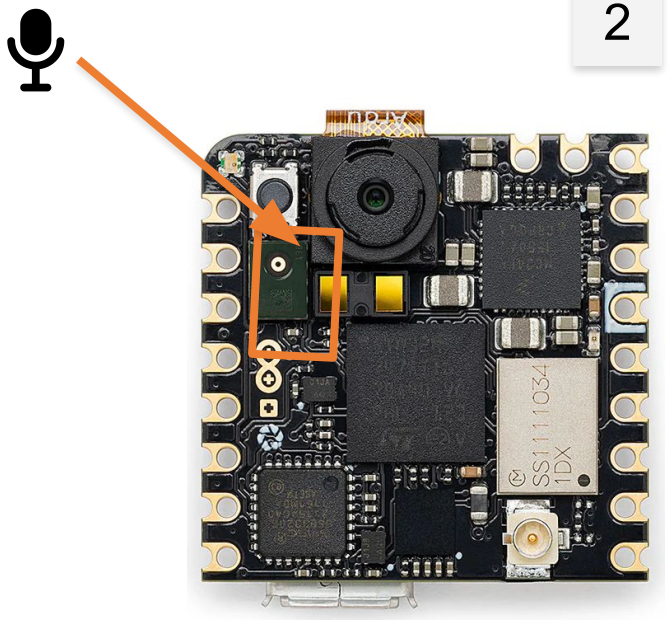
Security & Privacy

Battery & Memory

MEMORY

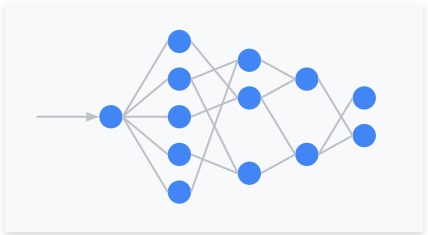
Run on resource constrained devices

Anatomy of a Keyword Spotting Application



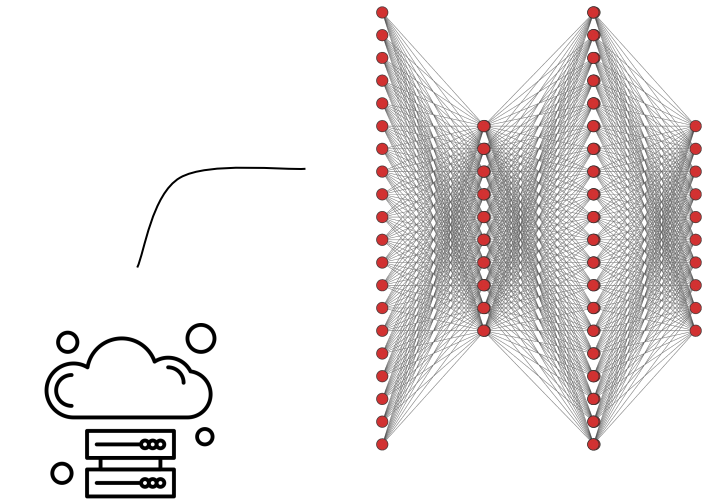
1 Continuously listen on the microcontroller

2 Process the data with **TinyML** at the edge

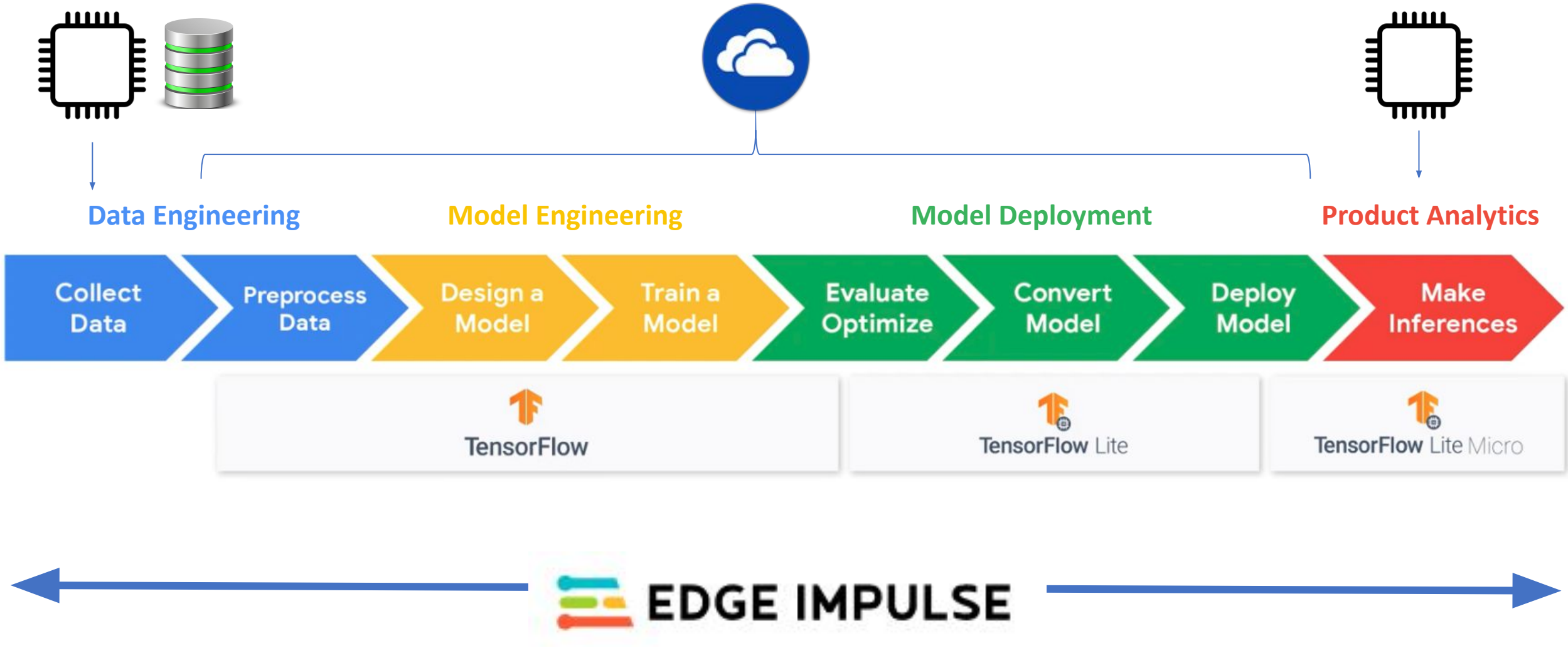


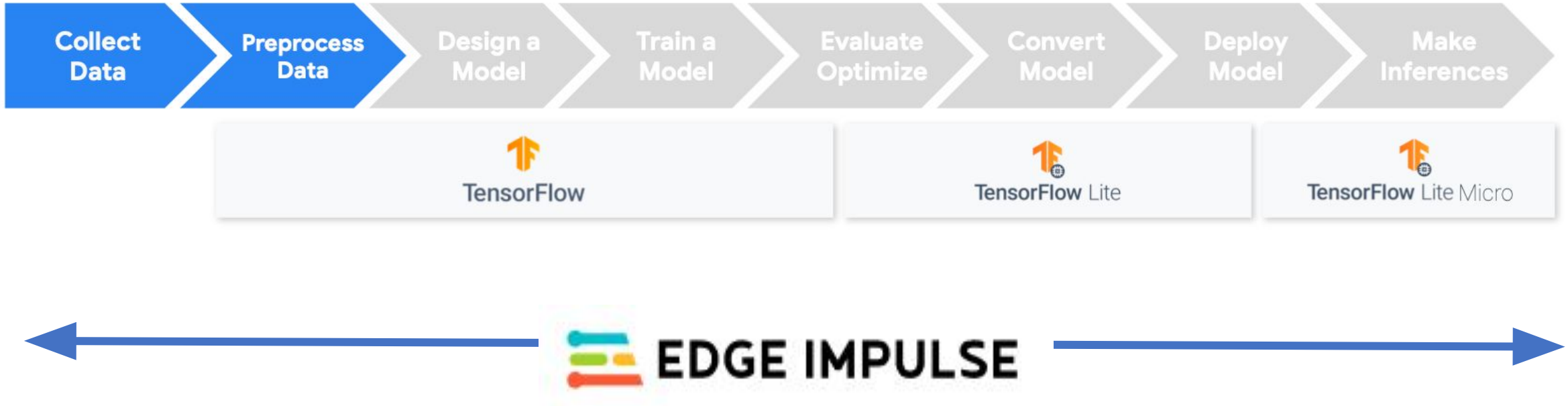
3 Send the data to the cloud when triggered

4 Process the full speech data with a large model in the cloud

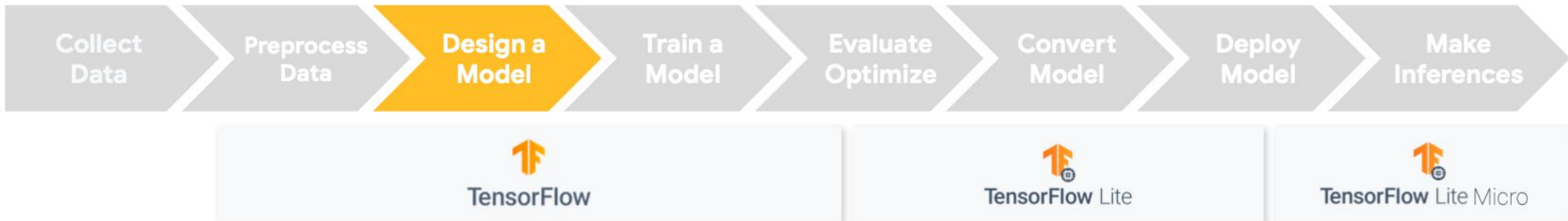


KWS Data Collection & Pre-Processing

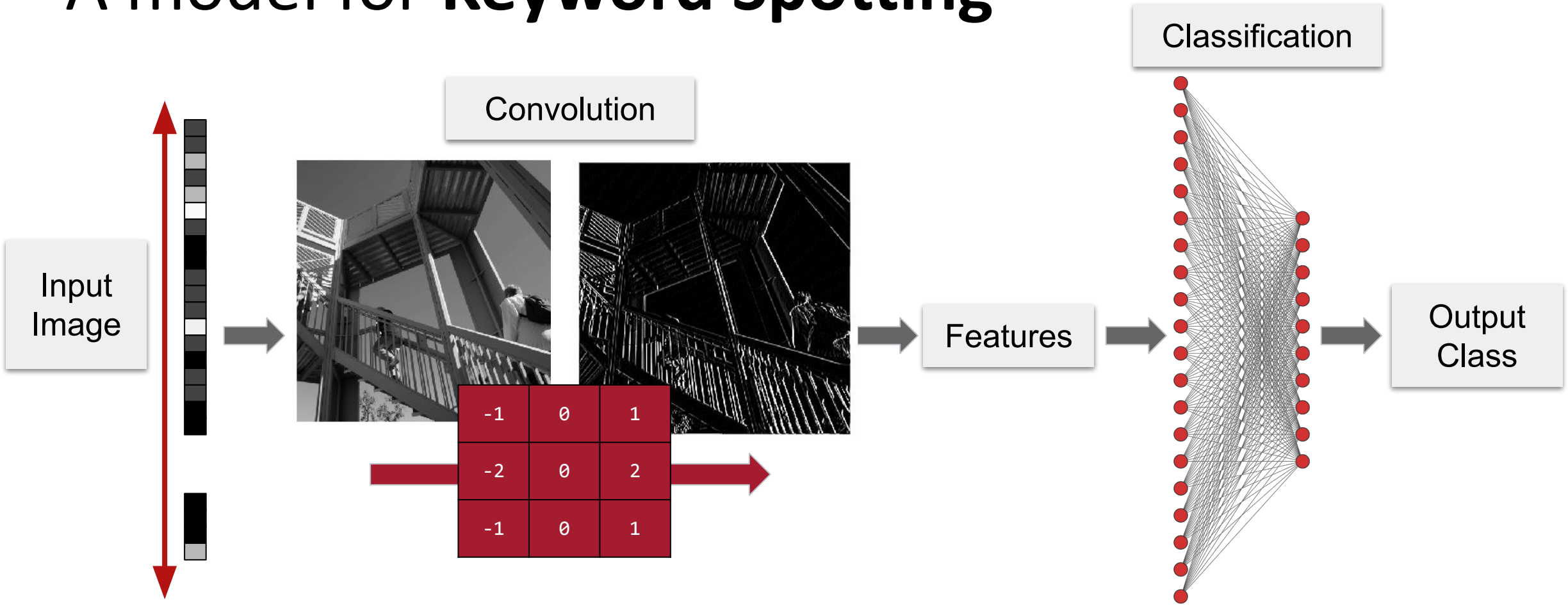

Data Engineering
Model Engineering
Model Deployment
Product Analytics
Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro
EDGE IMPULSE


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro
EDGE IMPULSE

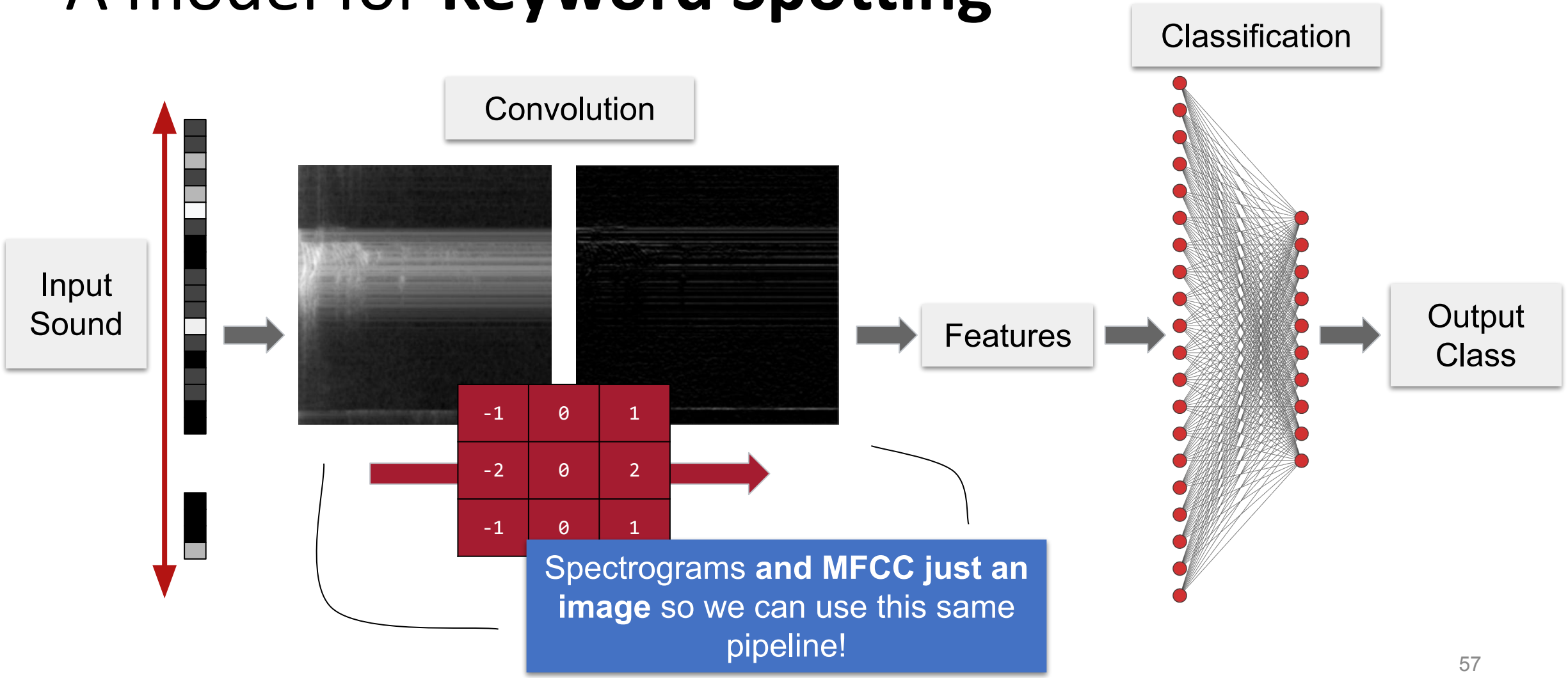
A Keyword Spotting Model


Collect Data
Preprocess Data
Design a Model
Train a Model
Evaluate Optimize
Convert Model
Deploy Model
Make Inferences
TensorFlow
TensorFlow Lite
TensorFlow Lite Micro

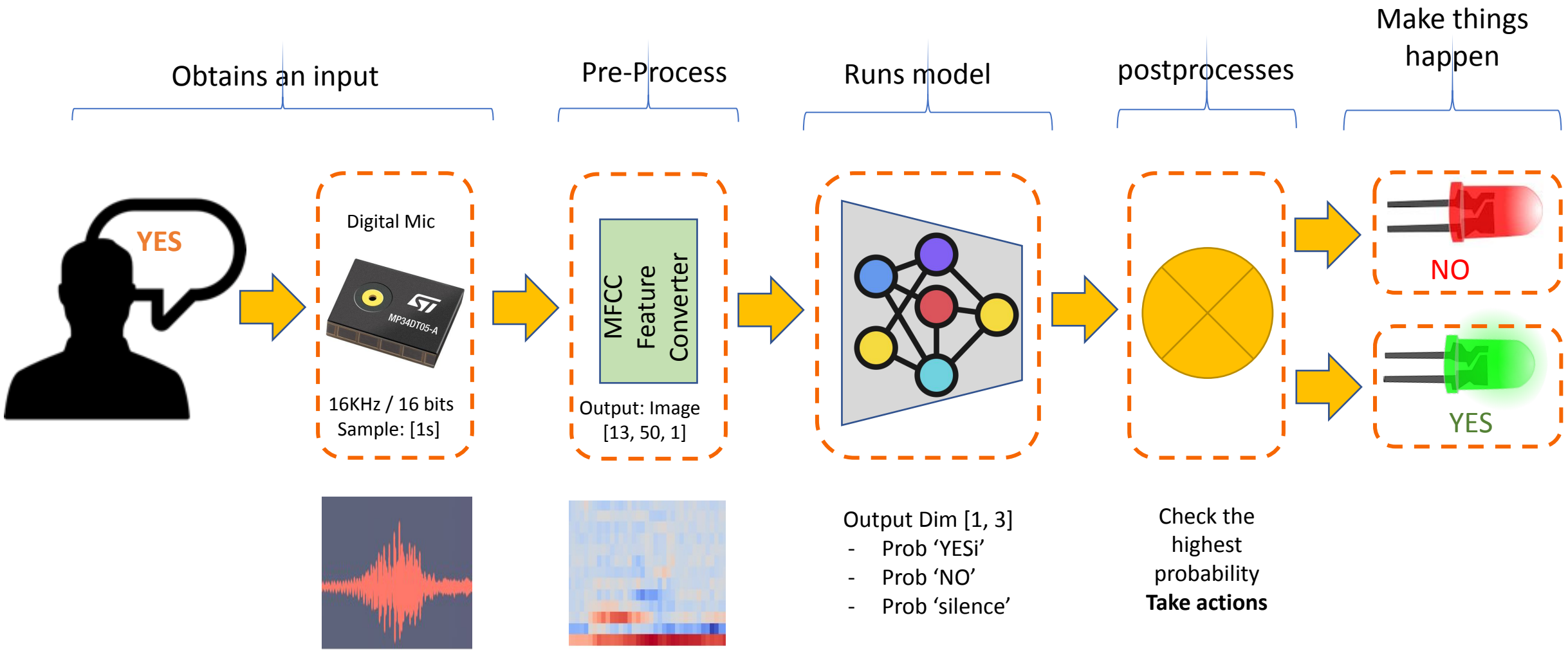
A model for **Keyword Spotting**



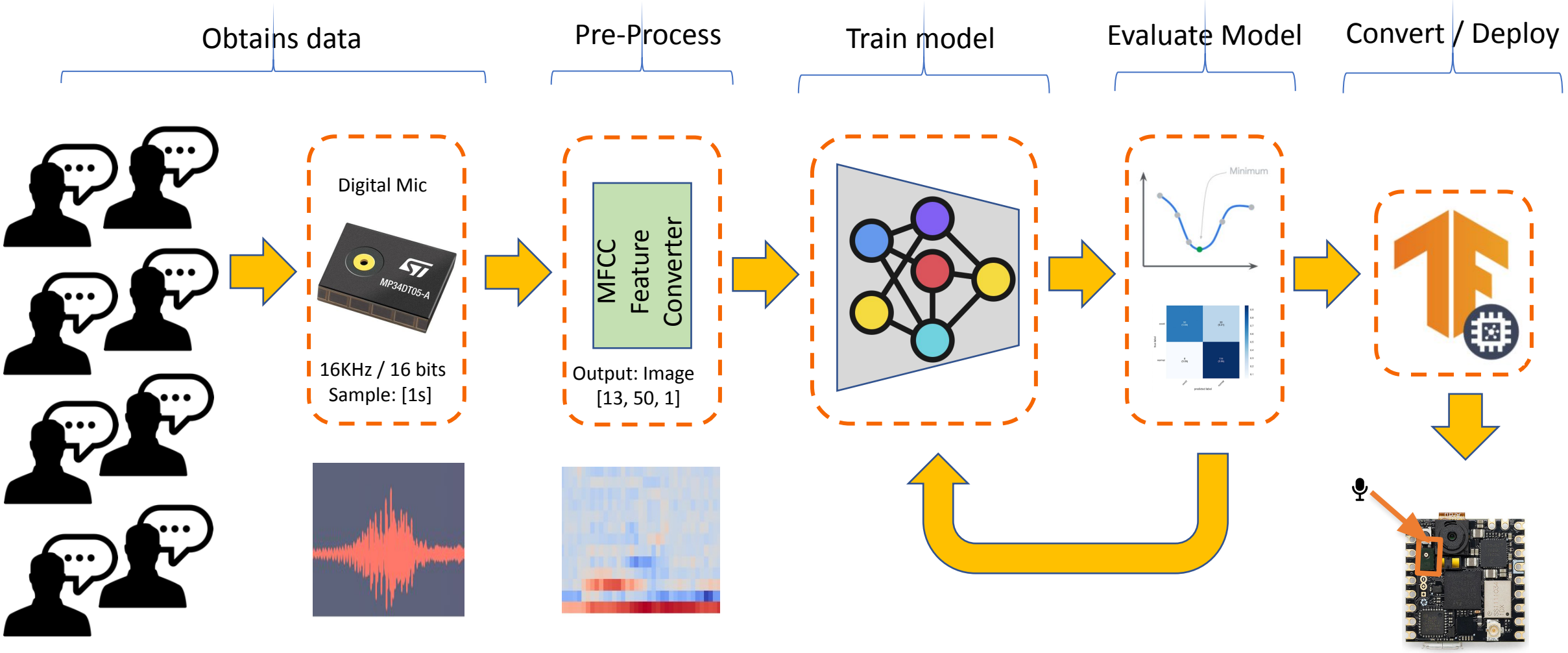
A model for **Keyword Spotting**



KeyWord Spotting (KWS) - Inference



KeyWord Spotting (KWS) – Create Model (Training)



Thanks

Workshop on
TinyML for
Sustainable Development

Prof. Jesús Alfonso López
jalopez@uao.edu.co
Universidad Autónoma de Occidente