EECE-4710 "IoT and TinyML"

KeyWord Spotting (KWS)

Cris Ababei

MARQUETTE UNIVERSITY

BE THE DIFFERENCE.

What is KeyWord Spotting?

Keyword Spotting v. General Speech Recognition

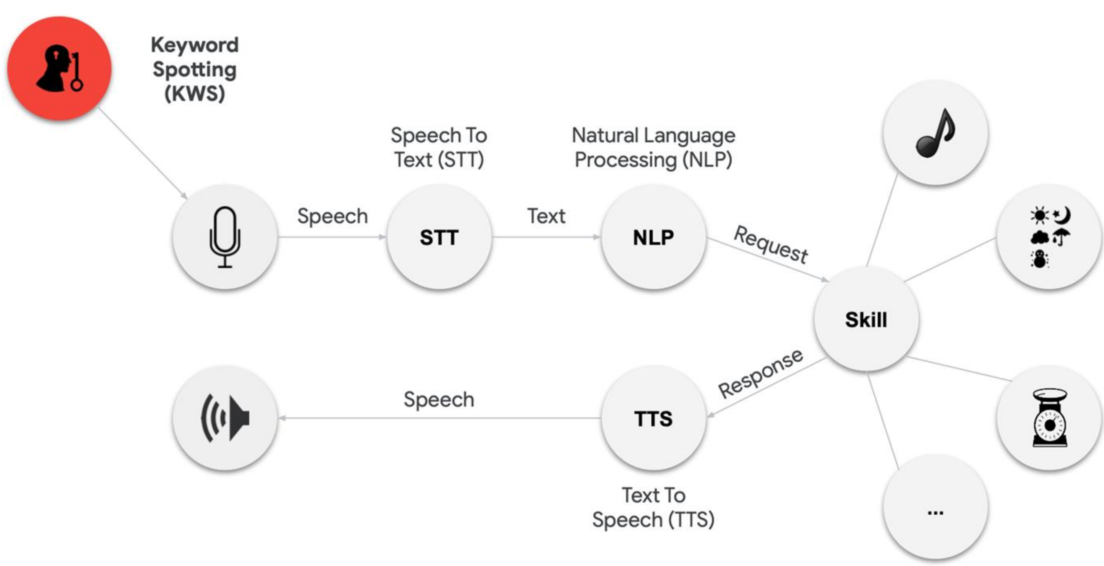
- **Keyword spotting** is one of the most successful examples of **TinyML**
  - Low-power, continuous, on-device
  - Common Voice SWTS* expands keyword spotting to more languages

    * **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

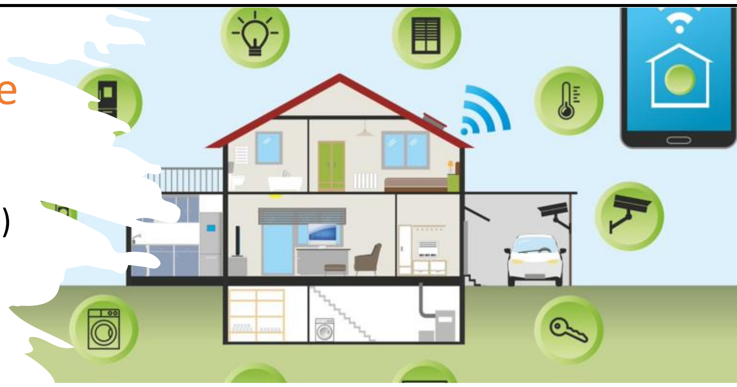
3



4

More than just voice

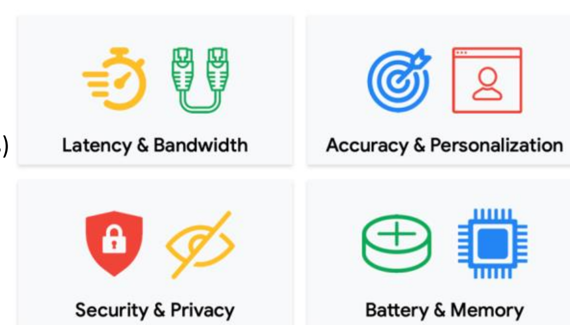
- Security (Broken Glass)
- Industry (Anomaly Detection)
- Medical (Snore)
- Nature (Bee, insect sound)



5

Challenges and Constraints

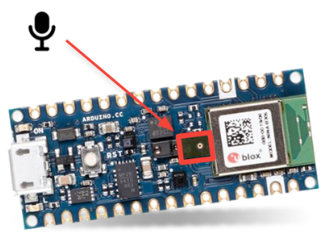
- **Latency**
  - Provide results quickly; respond in real-time to user
- **Bandwidth**
  - Minimize data sent over network (slow and expensive)
- **Accuracy**
  - Listen continuously, but only trigger at right time(s)
- **Personalization**
  - Trigger for user not background noise
- **Security & Privacy**
  - Safeguard data sent to cloud
- **Battery**
  - Limited energy, operate on coin-cell battery
- **Memory**
  - Run on resource constrained devices



6

6

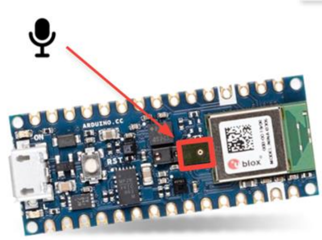
Anatomy of a Keyword Spotting Application



1 Continuously listen on the microcontroller

7

Anatomy of a Keyword Spotting Application

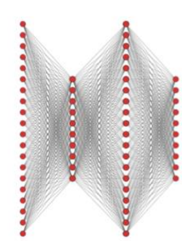


1 Continuously listen on the microcontroller

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

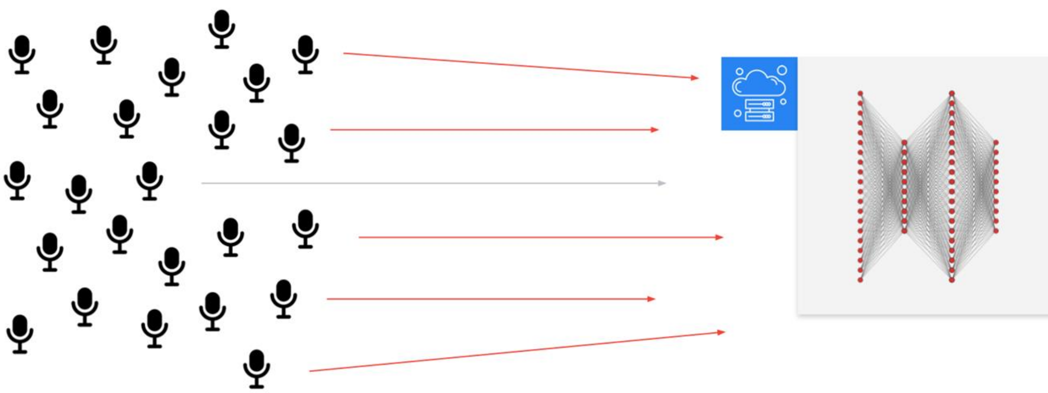
3 Send the data to the cloud when triggered

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



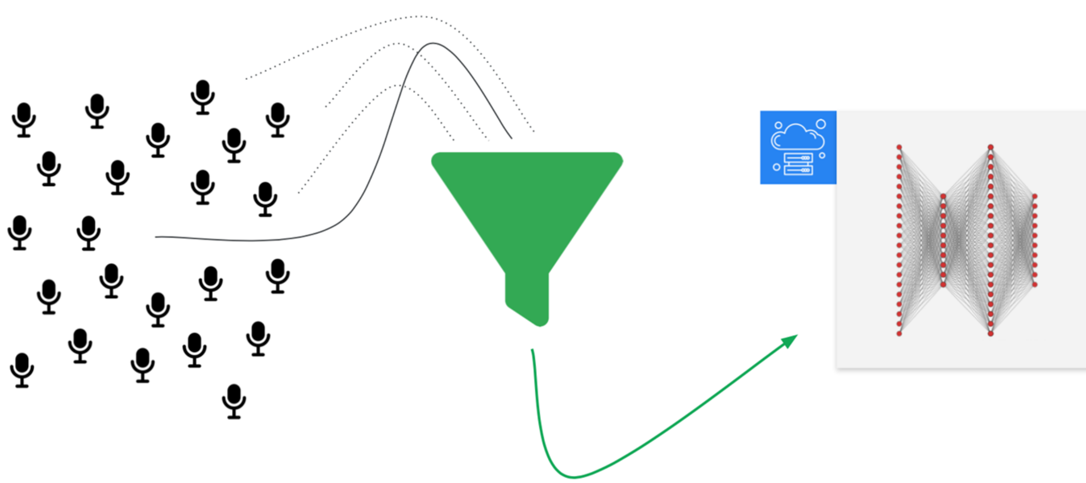
8

Anatomy of a Keyword Spotting Application



9

Anatomy of a Keyword Spotting Application



10

Keyword Spotting
Datasets

11

11

How do we build a good dataset?

- Who are the **users**?
- What do they **need**?
- What **task** are they trying to solve?
- How do they **interact** with the system?
- How does the **real world** make this hard?

12

Speech Commands: A Dataset for Limited-Vocabulary Speech Recognition

Pete Warden
Google Brain
Mountain View, California
petewarden@google.com

April 2018

https://arxiv.org/pdf/1804.03209.pdf

13

Requirements

"yes"
"no"

Common Use

"left"
"right"
"go"
"stop"

Robotics

"one"
"two"
"four"
"six"

Numbers

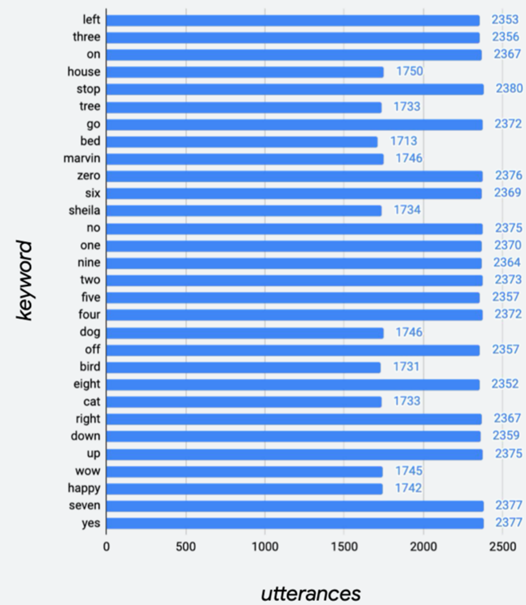
V1: 10 words
V2: 35 words

14

Data Collection

- **2,618** volunteers
    - consented to have their voices redistributed
    - Variety of accents
- > 1,000 examples for **each** keyword
- **Browser-based**

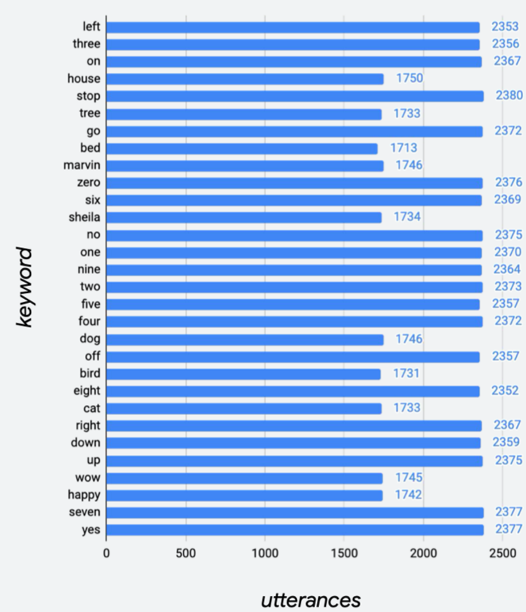
    (no app to install)



15

Data Validation

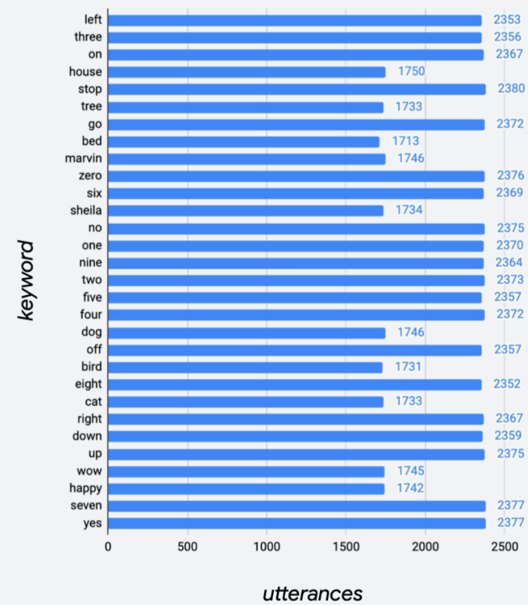
- Some data is **unusable**
    - Too quiet, wrong word, etc
- Started with **automated tools**
    - Remove low volume recordings
    - Extract loudest 1s (from 1.5sec examples)
- All 105,829 remaining utterances **manually reviewed** through crowdsourcing



16

Sustaining KWS Research

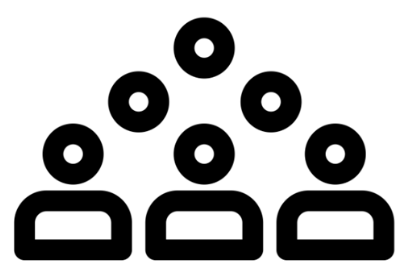
- Speech Commands is now in **v2**
  - **Expanded to 35** keywords from original 10
- Includes train/validation/test splits
- Expand to **new languages**?



17

Common Voice

- **Crowdsourcing** platform



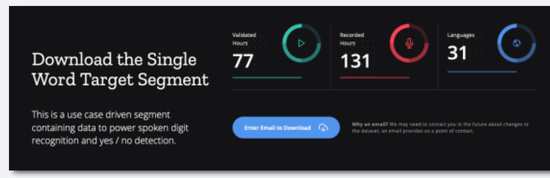
https://commonvoice.mozilla.org/en

18

Single Word Target Segment

A ***speech commands-style*** dataset for **18 languages**

- “Yes” // “no”
- “hey” & “Firefox”
- **digits** 0-9



https://commonvoice.mozilla.org/en/datasets

19

Food for Thought

QC (Quality Control)

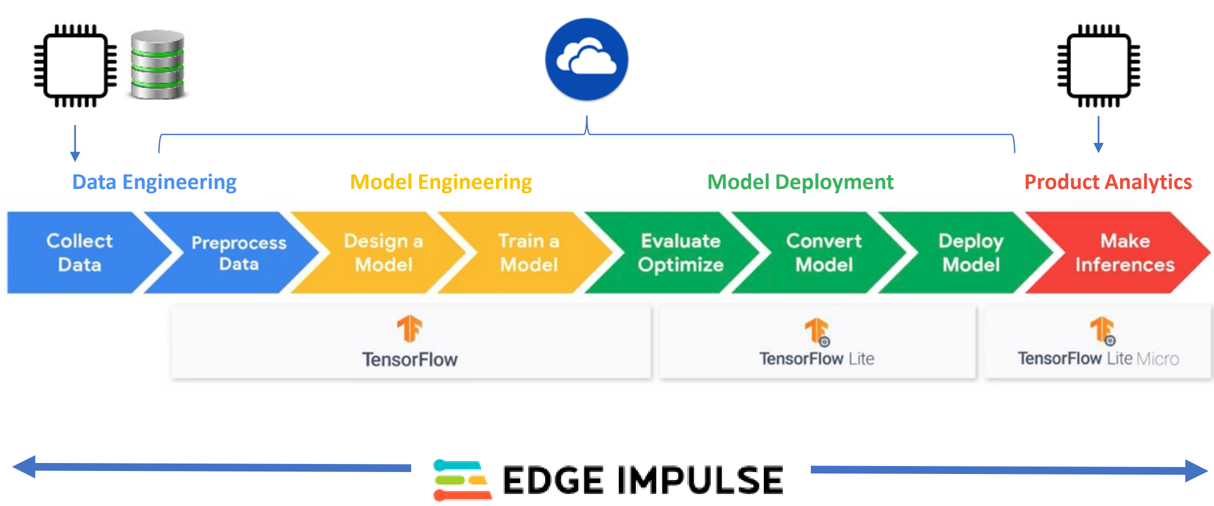
- Need to keep **only** what a human can hear
- Microphone issues
- **Noisy** backgrounds

20

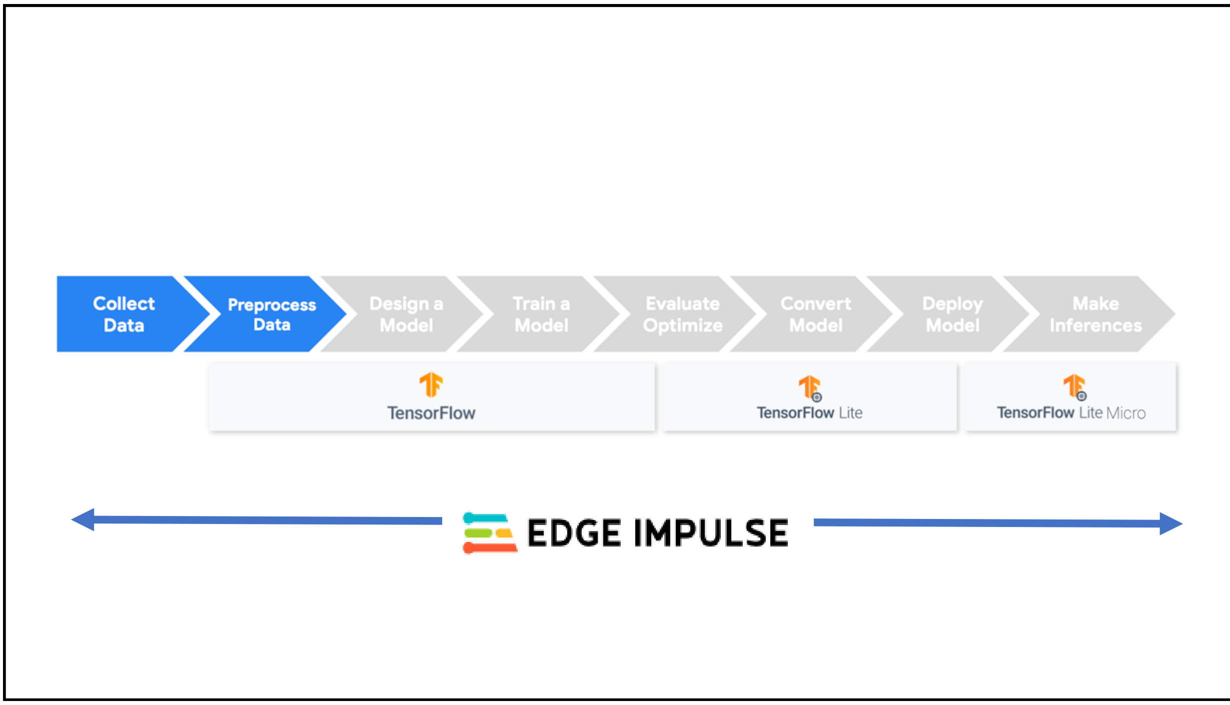
KWS Data Collection & Pre-Processing

21

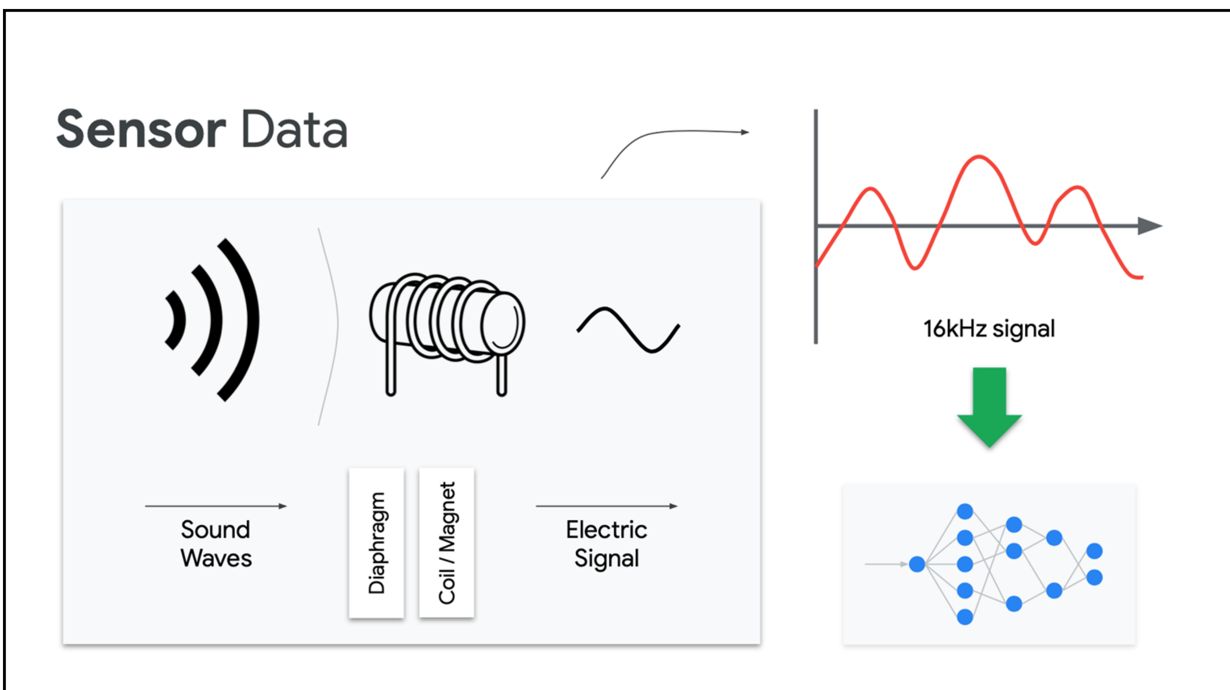
21



22


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

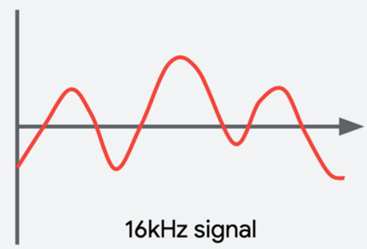
23


Sensor Data
Sound Waves
Diaphragm
Coil / Magnet
Electric Signal
16kHz signal

24

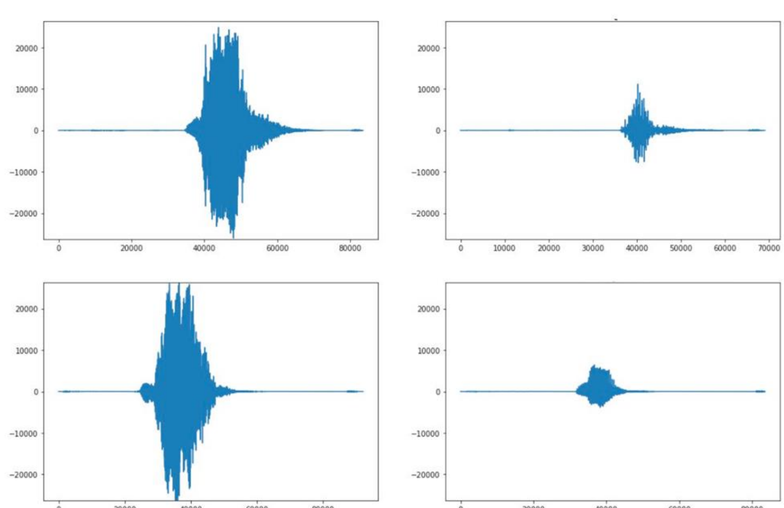
Sensor Data

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



25

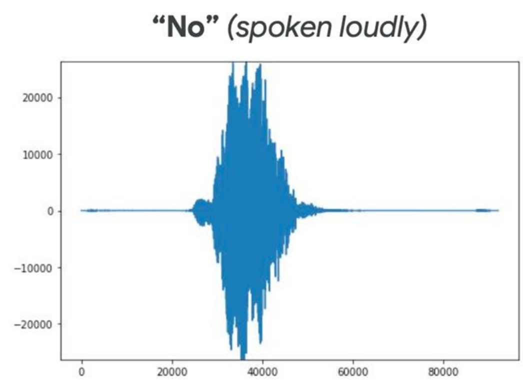
Guess!



26

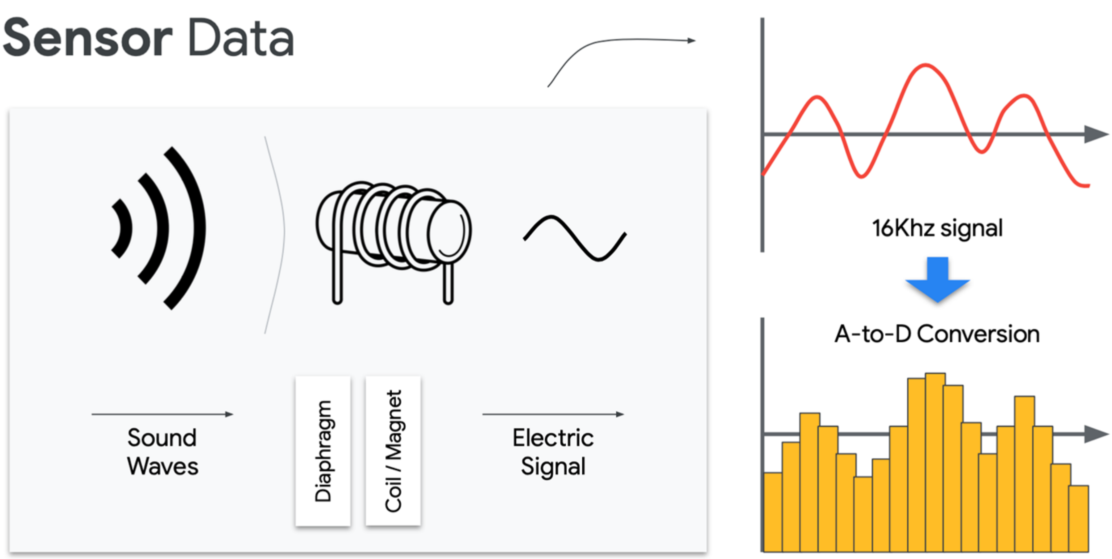
What are interesting **challenges**?

- It is a continuous signal, so **when does the word start**?
- How do you **“align”** on the starting point?
- How do we **extract the vital parts** of the signal that matter?



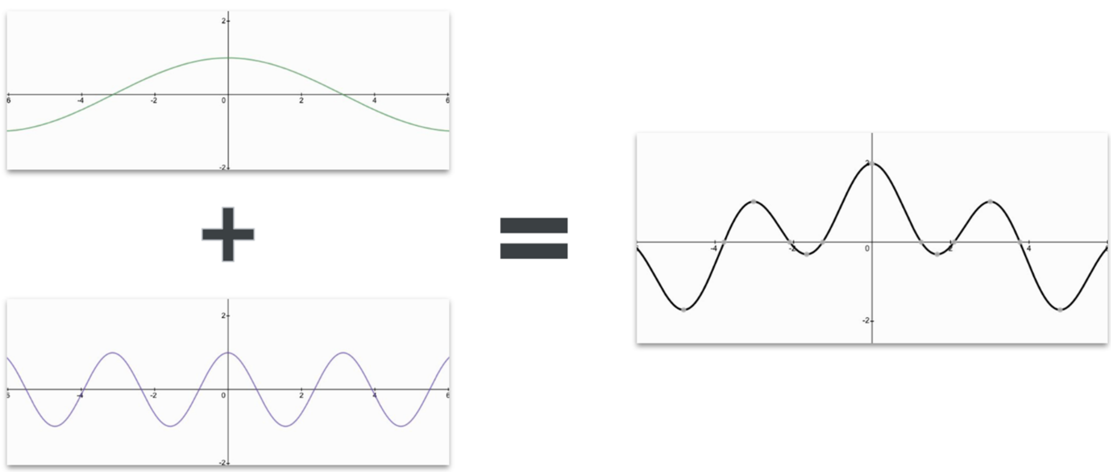
27

Sensor Data



28

Signal **Components**?

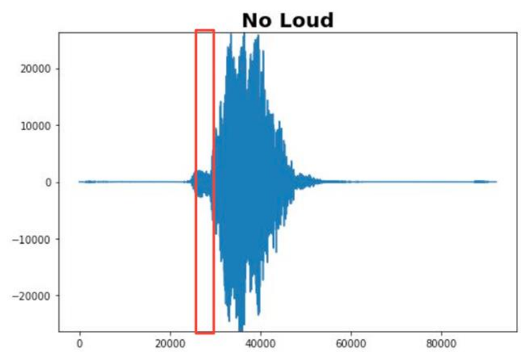


29

Signal **Components**?

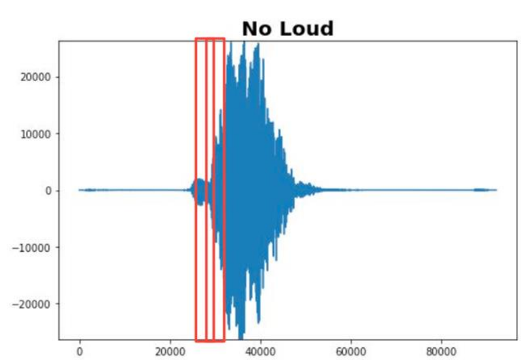
30

Data Preprocessing



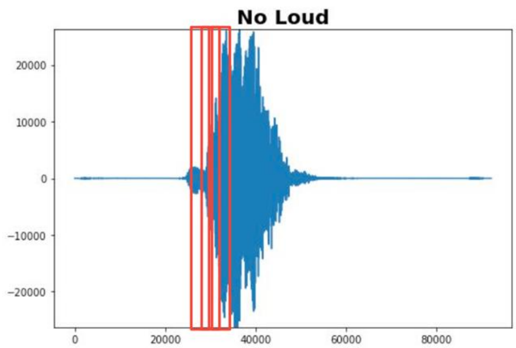
31

Data Preprocessing



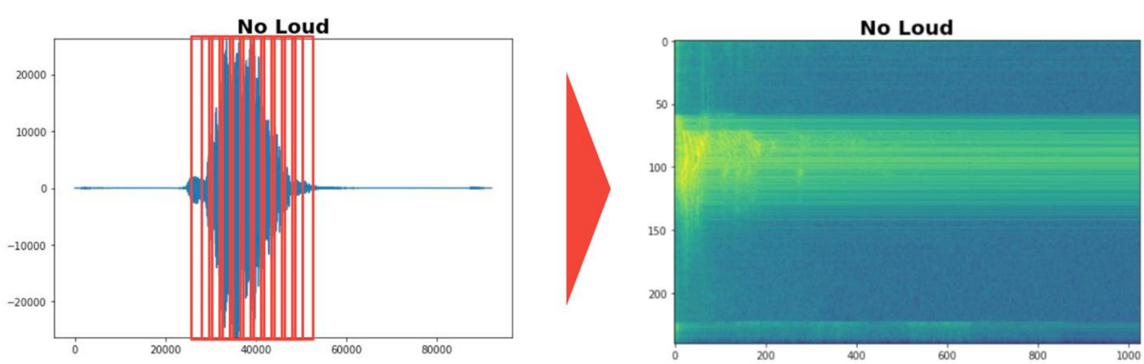
32

Data Preprocessing



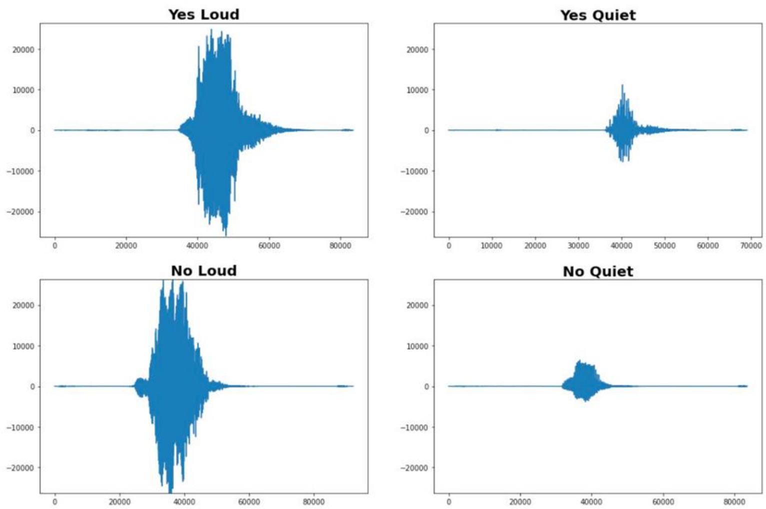
33

Data Preprocessing: **Spectrograms**



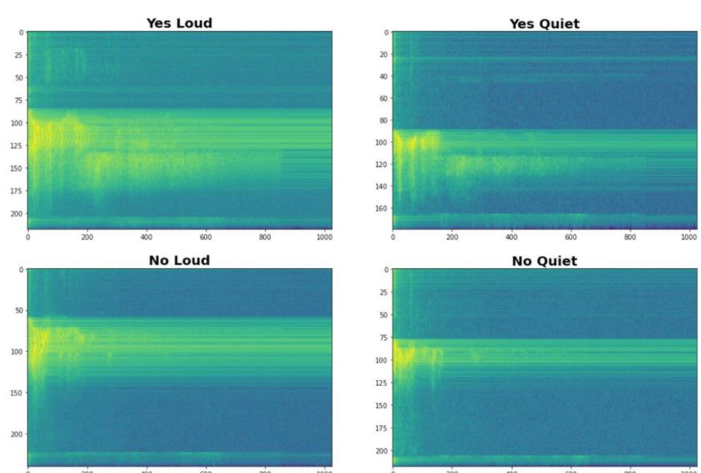
34

Data Preprocessing: **Spectrograms**



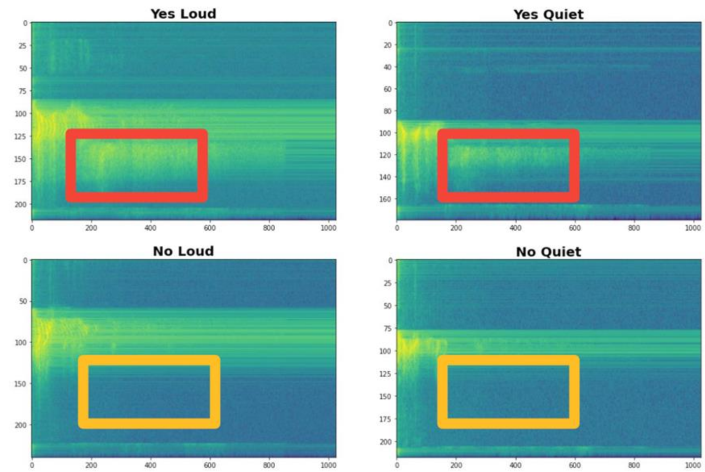
35

Data Preprocessing: **Spectrograms**



36

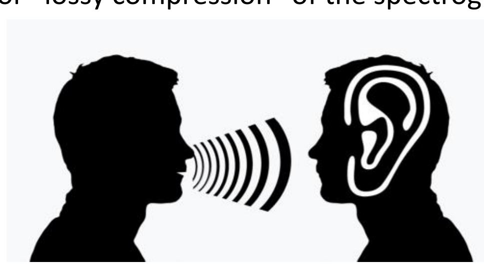
Data Preprocessing: **Spectrograms**



37

Drawbacks of the Spectrogram

1. We hear/perceive pitch exponentially in frequency - because freq. is exponential of our perception f = 440*2^(p/12). So, we do not want to include as many bins from high frequencies, because we would not be able to make much of a difference between bins at high freq.
2. We perceive intensity logarithmically in loudness.
3. Spectrograms have a lot of freq. bins; probably more than we need. So, we want to do a “dimension reduction” or “lossy compression” of the spectrogram that hopefully retains important aspects.

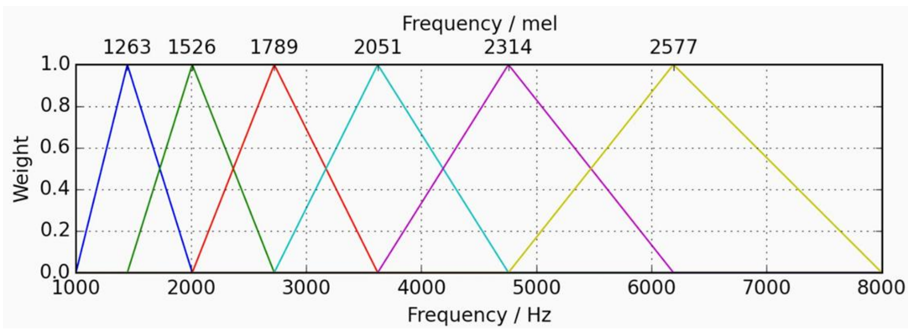


The lower band frequencies are much crisper to us

38

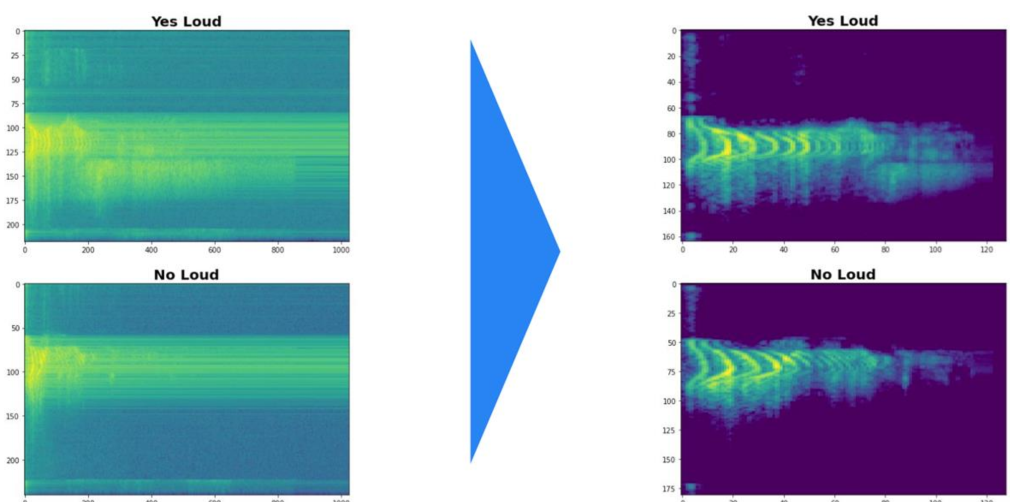
Mel (Triangle) Filterbanks

- Take freq. bins nonlinearly to match our perception
- Take say 40 bins (each bin is a triangle) of Mel Filterbanks and apply (multiply) with the Spectrogram to get: Mel Spectrogram



39

Spectrograms vs. Mel Spectrograms

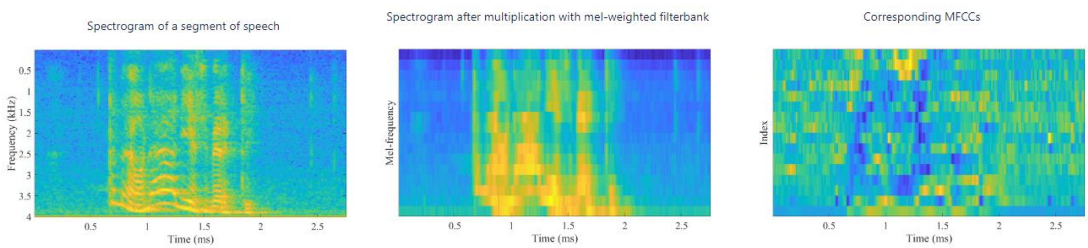


40

40

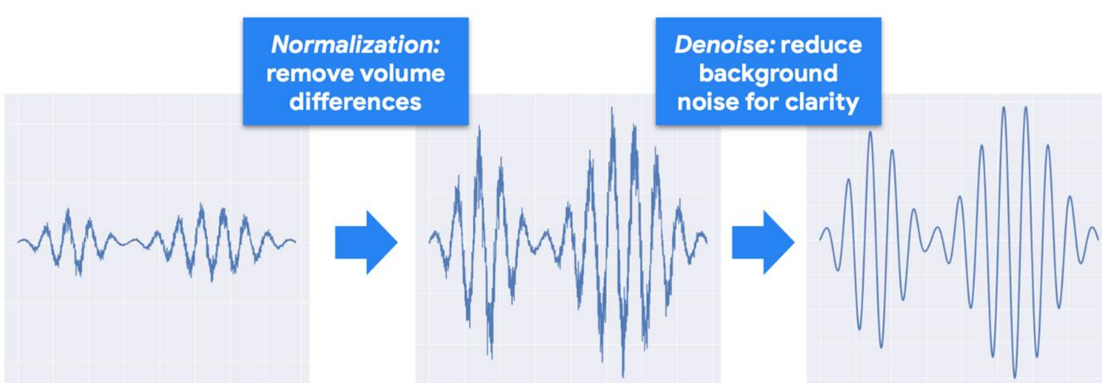
Mel-Frequency Cepstral Coefficients (MFCCs)

- Mel Frequency Cepstral Coefficients (MFCCs) are a feature widely used in automatic speech and speaker recognition; concisely describe the overall shape of a spectral envelope.
- How to calculate MFCCs
  1. Frame the signal into short frames.
  2. For each frame calculate the periodogram estimate of the power spectrum.
  3. Apply the Mel filterbank to the power spectra, sum the energy in each filter.
  4. Take the logarithm of all filterbank energies.
  5. Take the DCT of the log filterbank energies.
  6. Keep DCT coefficients 2-13, discard the rest.



41

Additional **Feature Engineering**



42

Spectrograms and MFCCs
Code Time!

SpectrogramsMFCCs.ipynb

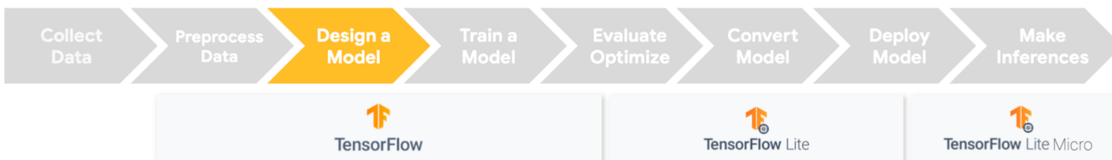
43

43

A Keyword Spotting
Model

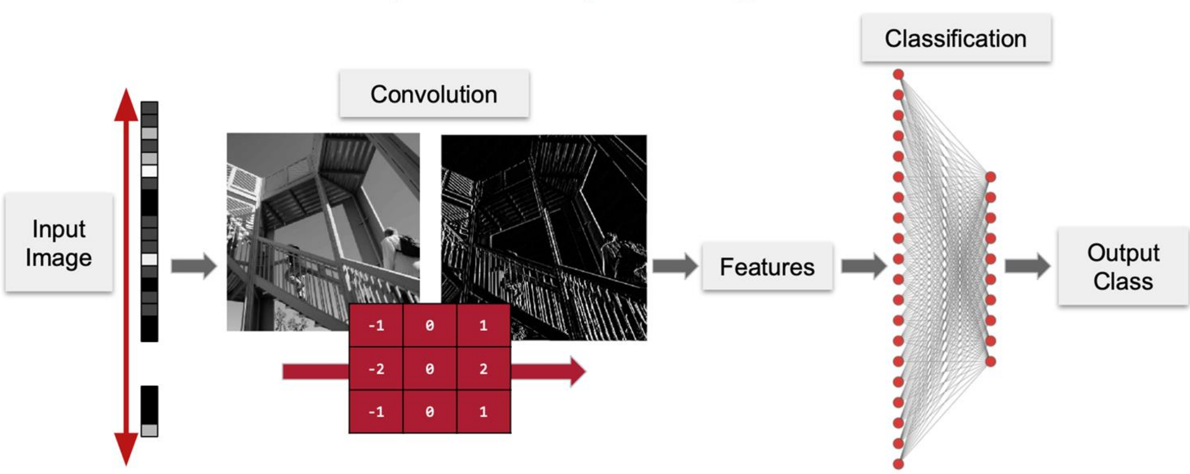
44

44



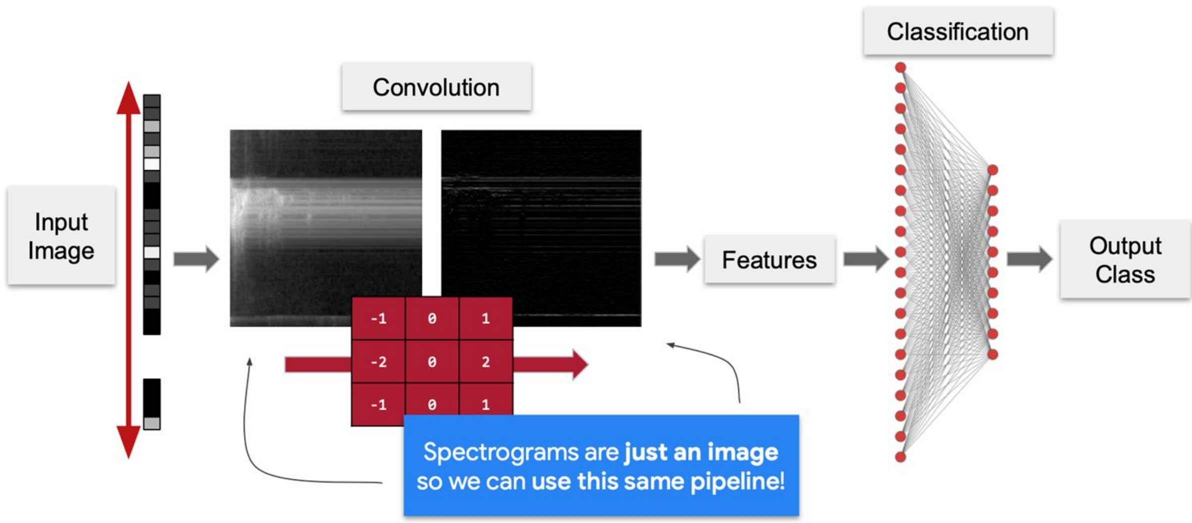
45

A model for **Keyword Spotting**



46

A model for **Keyword Spotting**



47

Credits

- A previous edition of this course was developed in collaboration with Dr. Susan C. Schneider of Marquette University.
- We are very grateful and thank all the following professors, researchers, and practitioners for jump-starting courses on TinyML and for sharing their teaching materials:
- Prof. Marcelo Rovai - TinyML - Machine Learning for Embedding Devices, UNIFEI
  - https://github.com/Mjrovai/UNIFEI-IESTI01-TinyML-2022.1
- Prof. Vijay Janapa Reddi - CS249r: Tiny Machine Learning, Applied Machine Learning on Embedded IoT Devices, Harvard
  - https://sites.google.com/g.harvard.edu/tinyml/home
- Prof. Rahul Mangharam – ESE3600: Tiny Machine Learning, Univ. of Pennsylvania
  - https://tinyml.seas.upenn.edu/#
- Prof. Brian Plancher - Harvard CS249r: Tiny Machine Learning (TinyML), Barnard College, Columbia University
  - https://a2r-lab.org/courses/cs249r_tinyml/

48

48

References

- Additional references from where information and other teaching materials were gathered include:
- Applications & Deploy textbook: “TinyML” by Pete Warden, Daniel Situnayake
  - https://www.oreilly.com/library/view/tinyml/9781492052036/
- Deploy textbook “TinyML Cookbook” by Gian Marco Iodice
  - https://github.com/PacktPublishing/TinyML-Cookbook
- Jason Brownlee
  - https://machinelearningmastery.com/
- TinyMLedu
  - https://tinyml.seas.harvard.edu/
- Professional Certificate in Tiny Machine Learning (TinyML) – edX/Harvard
  - https://www.edx.org/professional-certificate/harvardx-tiny-machine-learning
- Introduction to Embedded Machine Learning - Coursera/Edge Impulse
  - https://www.coursera.org/learn/introduction-to-embedded-machine-learning
- Computer Vision with Embedded Machine Learning - Coursera/Edge Impulse
  - https://www.coursera.org/learn/computer-vision-with-embedded-machine-learning

49