Mobile and Sensor Systems

Lecture 5: Modeling and Inference

Dr. Sourav Bhattacharya

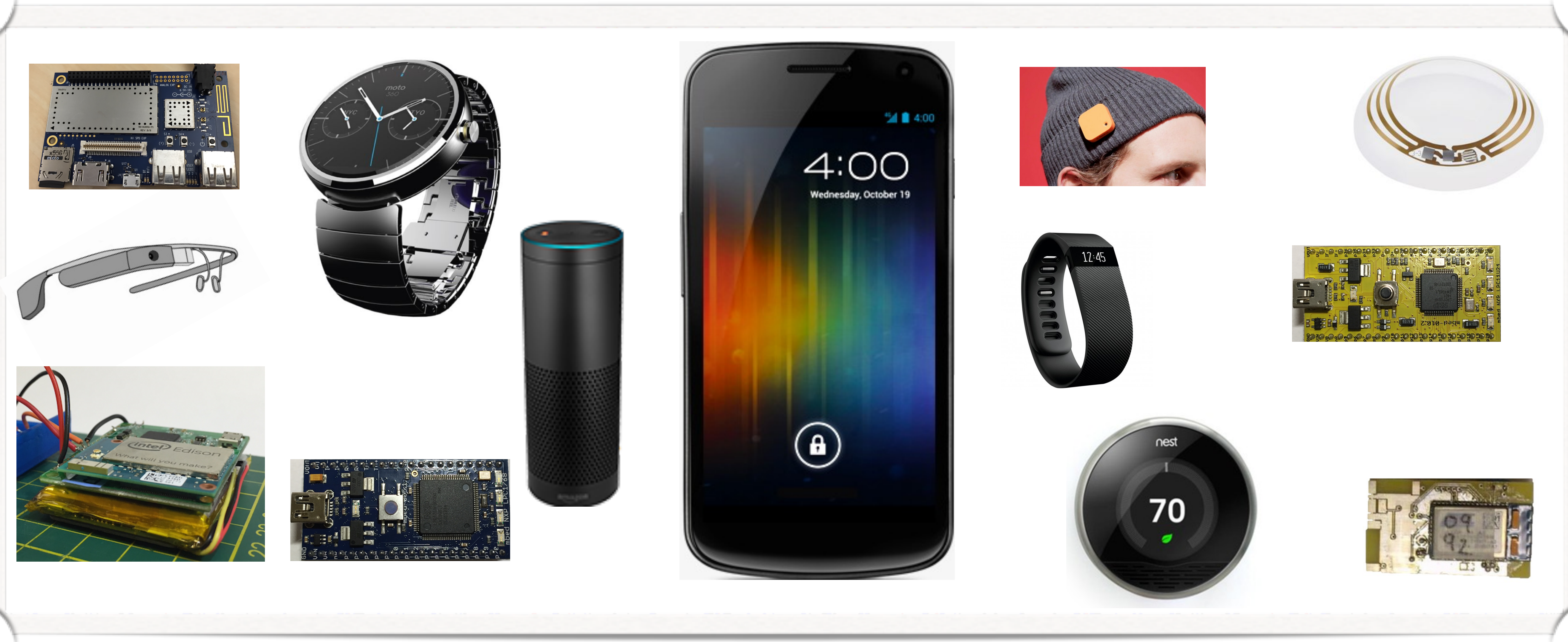
UNIVERSITY OF CAMBRIDGE

Lecture Overview

- Introduction to mobile and wearable sensing
- Mobile sensing applications
- Understanding the key tasks in mobile sensing
- Challenges in mobile sensing
- Case study: Modeling audio using Deep Neural Networks
- Open research questions

UNIVERSITY OF CAMBRIDGE

Mobile and Wearable Sensing



The mobile phone and wearable sensing domain is filled with **hacks**, and imaginative techniques that are used to circumvent the limitations of a platform that was **designed for a different purpose.**

Mobile / Wearable Sensing Vs. Sensor Networks

Mobile Sensing

- Well suited for human activities
- General purpose sensors, often not well suited for accurate sensing of the target phenomena
- Multi-tasking OS. Main purpose is to support various applications
- Low cost of deployment and maintenance (millions of users charge their devices)

Sensor Networks

- Well suited for sensing the environment
- Specialized sensors, designed to accurately monitor specific phenomena
- All resources dedicated to sensing
- High cost deployment and maintenance (regular charging thousands of sensor nodes)

Mobile Sensing Applications

Individual sensing:

- fitness applications
- behaviour intervention applications

Group/community sensing:

- groups to sense common activities and help achieving group goals
- examples: assessment of neighbourhood safety, environmental sensing, collective recycling efforts

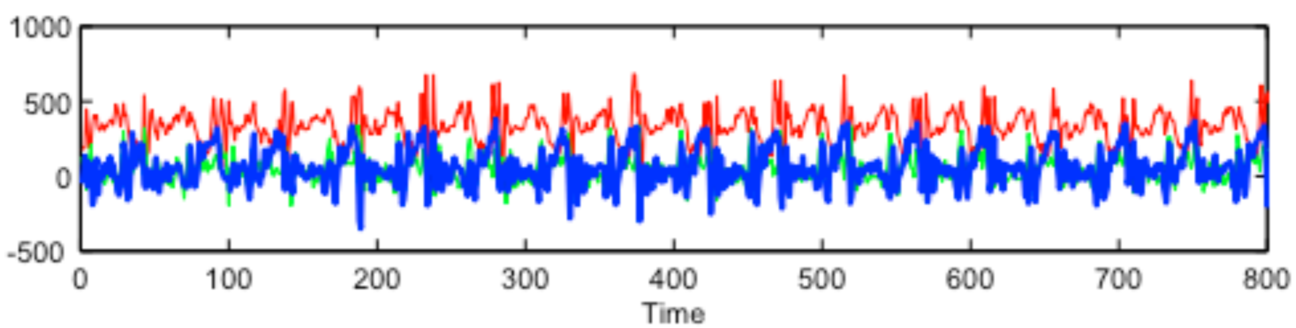
Urban-scale sensing:

- large scale sensing, where large number of people have the same application installed
- examples: tracking speed of disease across a city, congestion and pollution in a city

UNIVERSITY OF CAMBRIDGE

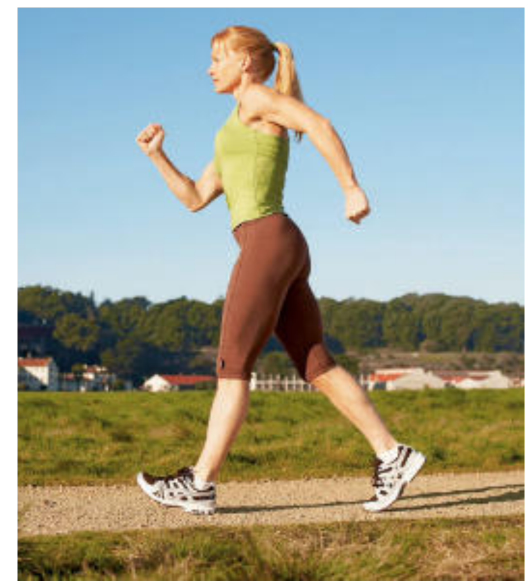
Human Activity Recognition

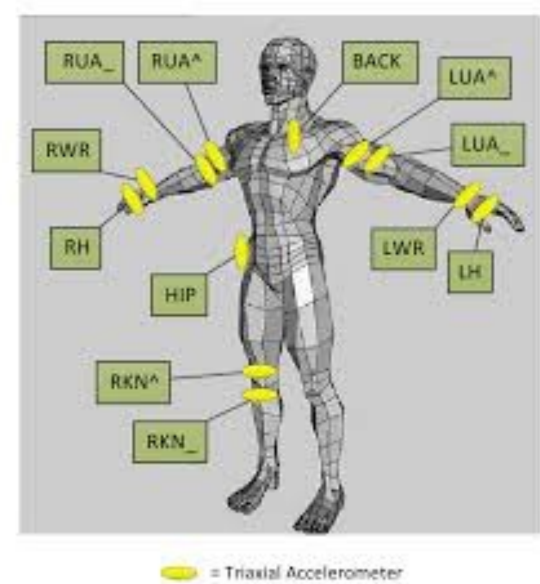
Sensor used:
- Accelerometer or Gyroscope



Example inference:
- Walking, running, biking, up/down stairs etc.

Applications:
- Health / behaviour intervention
- Fitness monitoring
- Sharing within a community

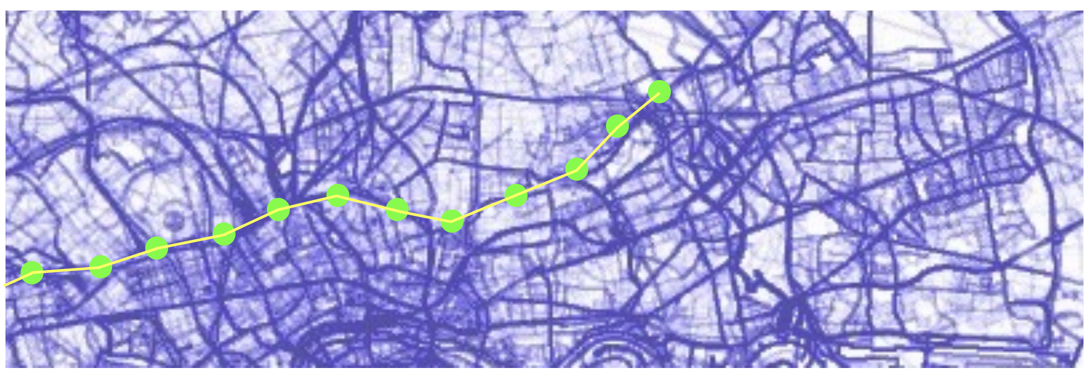




UNIVERSITY OF CAMBRIDGE

Transportation-mode Detection

Sensor used:
- Accelerometer or Gyroscope
- GPS, WiFi localization



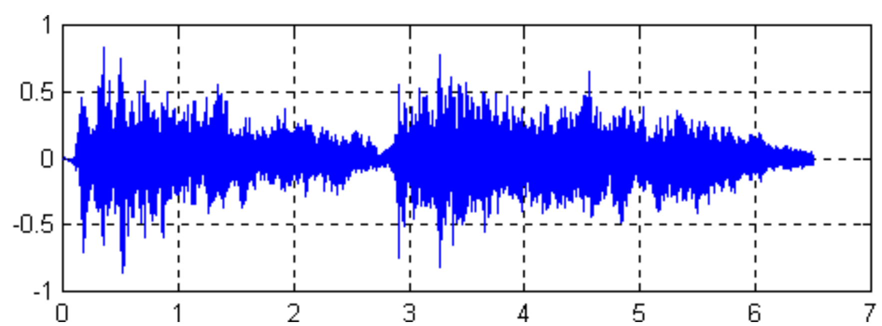
Example inference:
- Bus, bike, tram, train, car etc.

Applications:
- Intelligent transportation
- Smart commuting

UNIVERSITY OF CAMBRIDGE

Emotion Detection

Sensor used:
- Microphone, bluetooth
- GPS, WiFi localization
- Map speaking features to emotional state



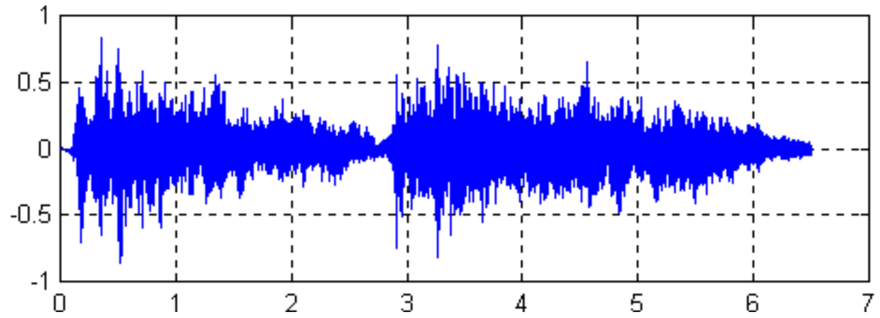
Example inference:
- Emotional state, location and co-location with others

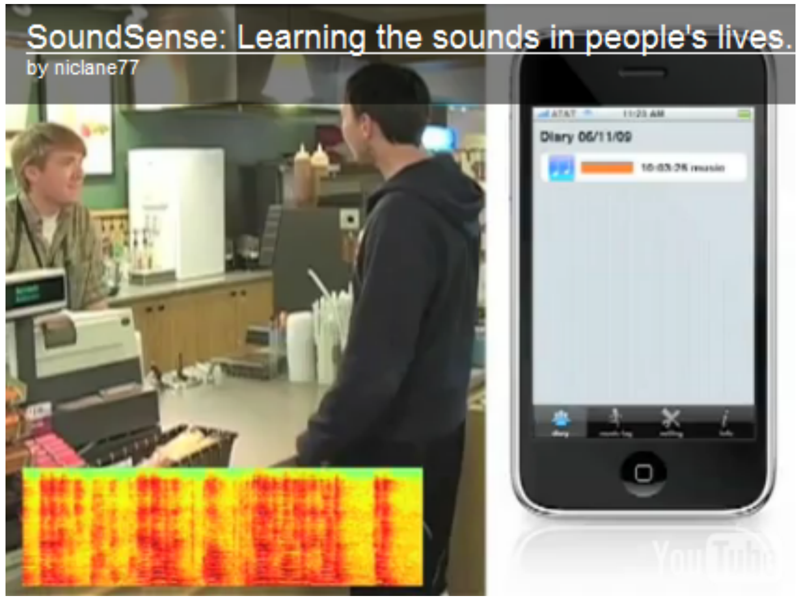
Applications:
- Behaviour intervention
- Computational social science
  - Using mobile sensing for quantifying theories in social science

UNIVERSITY OF CAMBRIDGE

Context and Environment

Sensor used:
- Microphone
- Camera





Example inference:
- Conversation, music, party, activity-related sound etc.

Applications:
- Automated diary
- Health and wellness

UNIVERSITY OF CAMBRIDGE

Challenges in Mobile Sensing

- Complex natural environment
- Heterogeneity of sensors
    - Vary in sampling frequency, sensitivity
- Noisy measurements
- Different sensor position and orientation
- Diverse population
- Privacy
- Limited processing and battery power

Common sensing platforms

Noisy data

Diverse user population

UNIVERSITY OF CAMBRIDGE

Challenges in Mobile Sensing

- **Sensing is resource intensive**

Battery

Memory

CPU

GPU

Storage

- **The purpose of the embedded platform is to support multiple applications**

- **A sensing application needs to maintain a balance between**
  - The amount of resource needed to operate
  - The accuracy of the detection that is achieved

UNIVERSITY OF CAMBRIDGE

Context Recognition: Machine Learning

Supervised Learning:

- Labeled data (training data)
- Objective: Learn a function from training data

$$\mathcal{F} : \mathbf{X} \rightarrow \mathbf{Y} \qquad \boldsymbol{x_i} \in \mathbb{R}^d$$

Feature vector	Label
$\boldsymbol{x}_1$	y_1
$\boldsymbol{x}_2$	y_2
$\vdots$	$\vdots$
$\boldsymbol{x}_n$	y_n

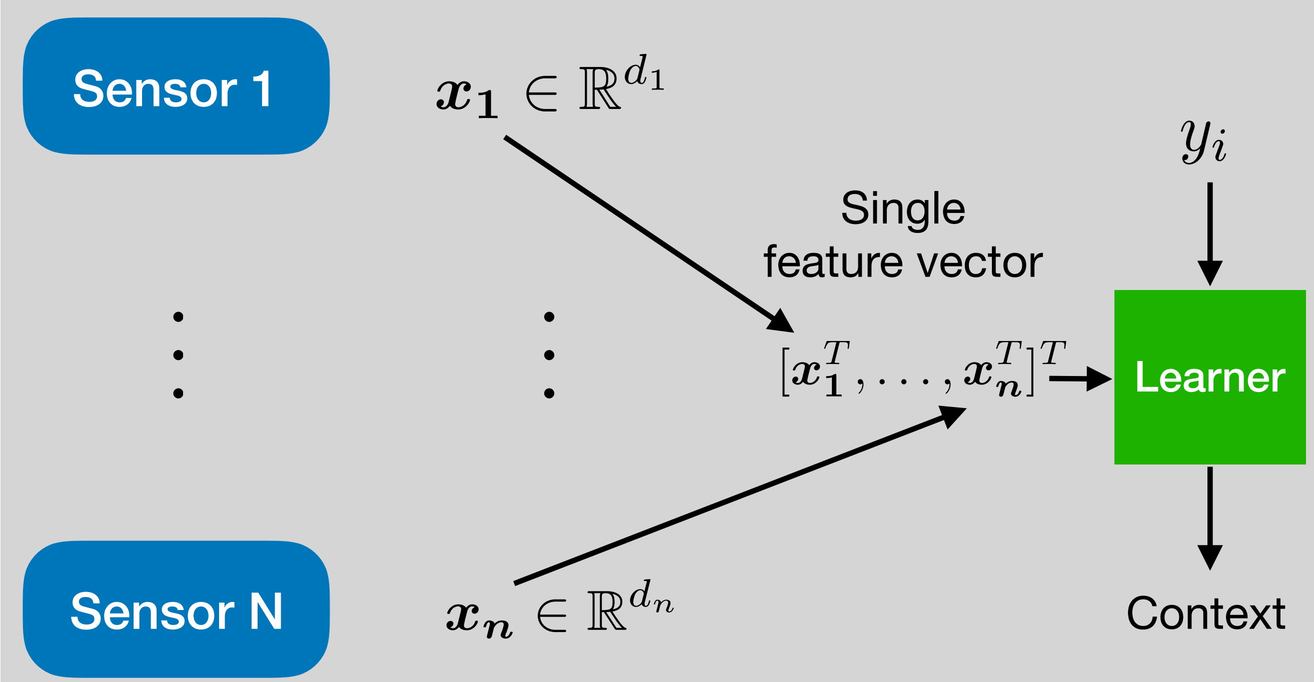
Classification

- Label is discrete / categorical variable

Regression

- Label is real-valued / continuous variable

In mobile sensing we have a large number of sensors



UNIVERSITY OF CAMBRIDGE

Context Recognition: Machine Learning

Supervised Learning:

- Labeled data (training data)
- Objective: Learn a function from training data

$$\mathcal{F} : \mathbf{X} \to \mathbf{Y} \qquad \boldsymbol{x_i} \in \mathbb{R}^d$$

Feature vector	Label
$\boldsymbol{x}_1$	y_1
$\boldsymbol{x}_2$	y_2
$\vdots$	$\vdots$
$\boldsymbol{x}_n$	y_n

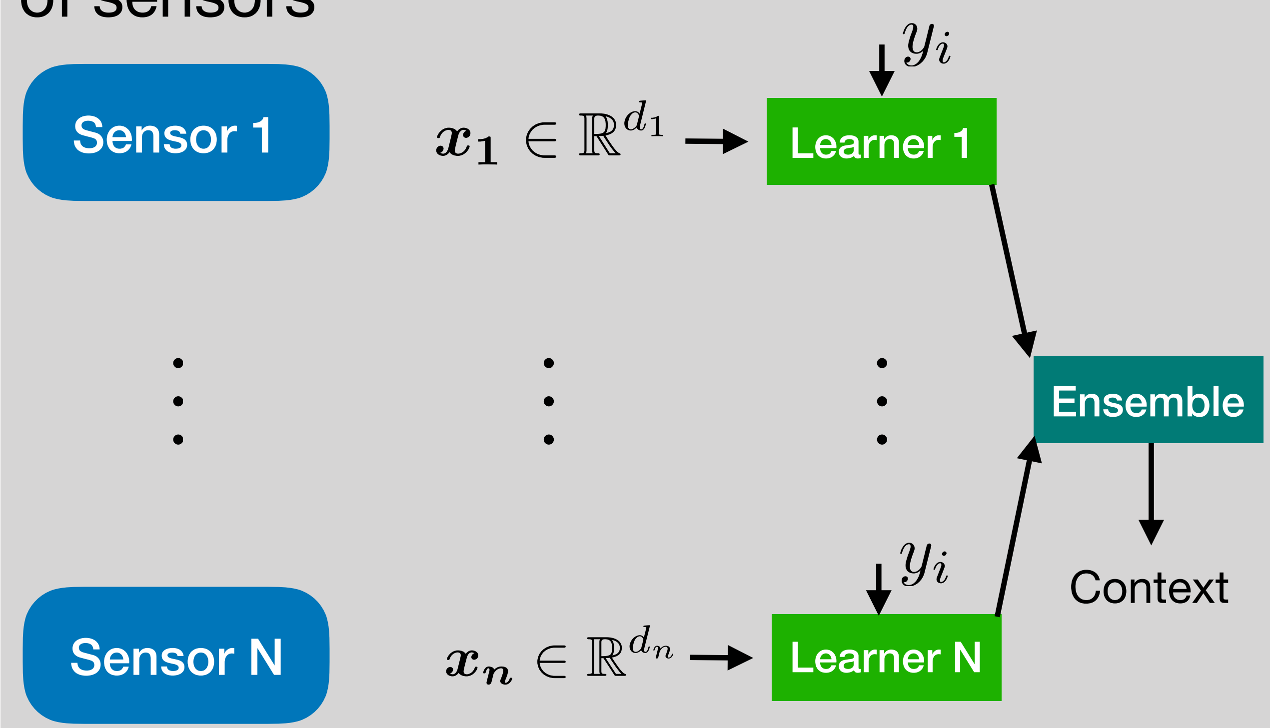
Classification

- Label is discrete / categorical variable

Regression

- Label is real-valued / continuous variable

In mobile sensing we have a large number of sensors

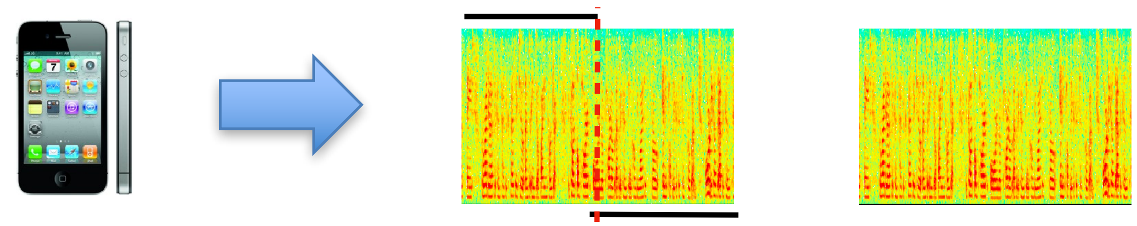


UNIVERSITY OF CAMBRIDGE

Development Design Pattern

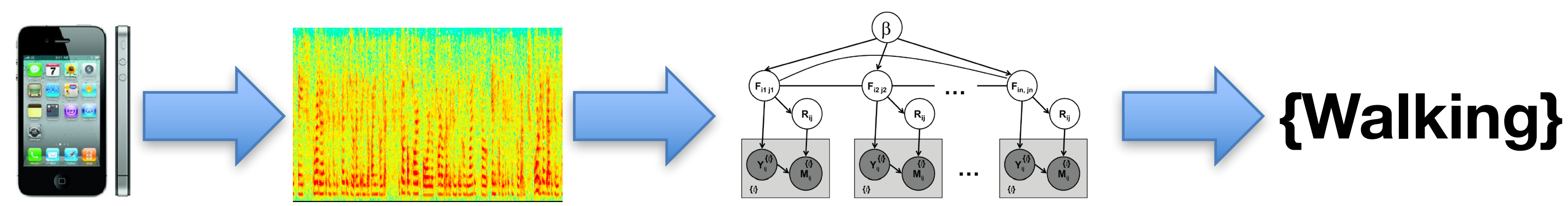
Data collection:

- Labeled or
- Unlabeled



Inference pipeline:

- Sensing
- Feature extraction
- Classification



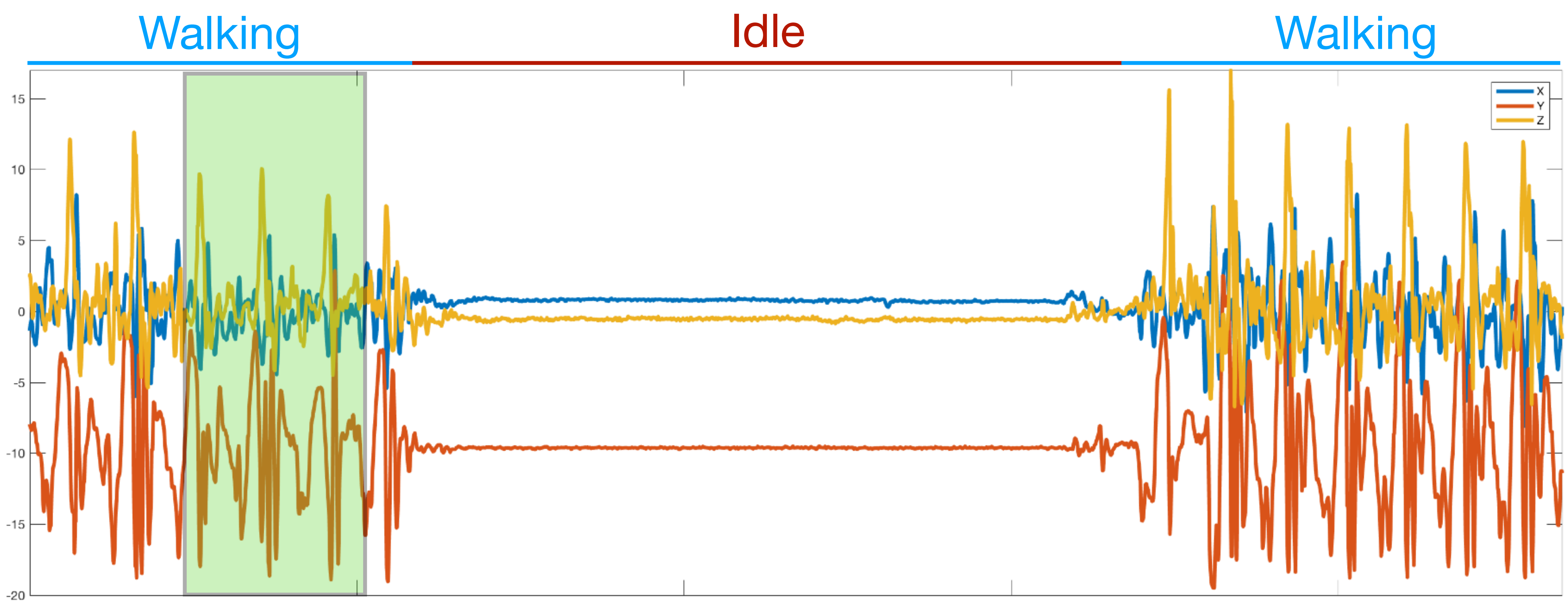
Sensing **Feature extraction** **Classification**

Mobile sensing app:

- Storage
- Networking
- Sharing, privacy

UNIVERSITY OF CAMBRIDGE

Case Study: Physical Activity Recognition



Accelerometer data

Feature engineering:

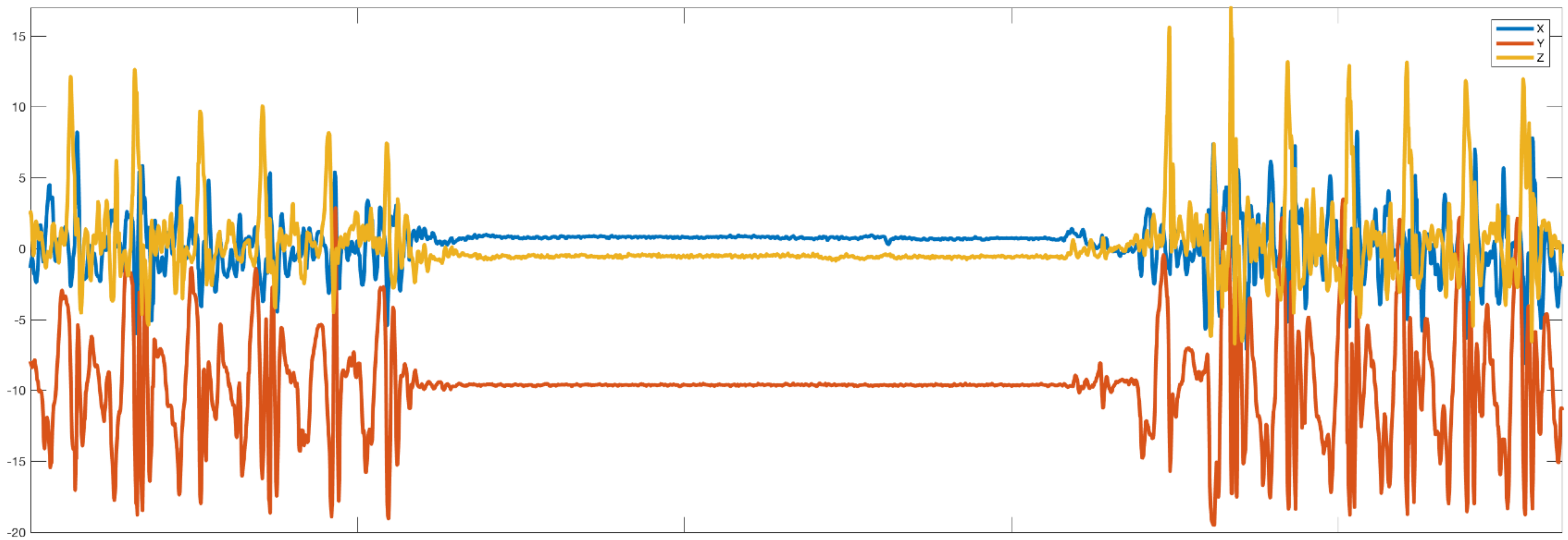
- Mean, variance, skewness, mean-crossing rate, peak etc.
- FFT, frequency bands, energy etc.

Supervised learning:

- Decision tree (C4.5)
- SVM
- Random forest

UNIVERSITY OF CAMBRIDGE

Continuous Sensing



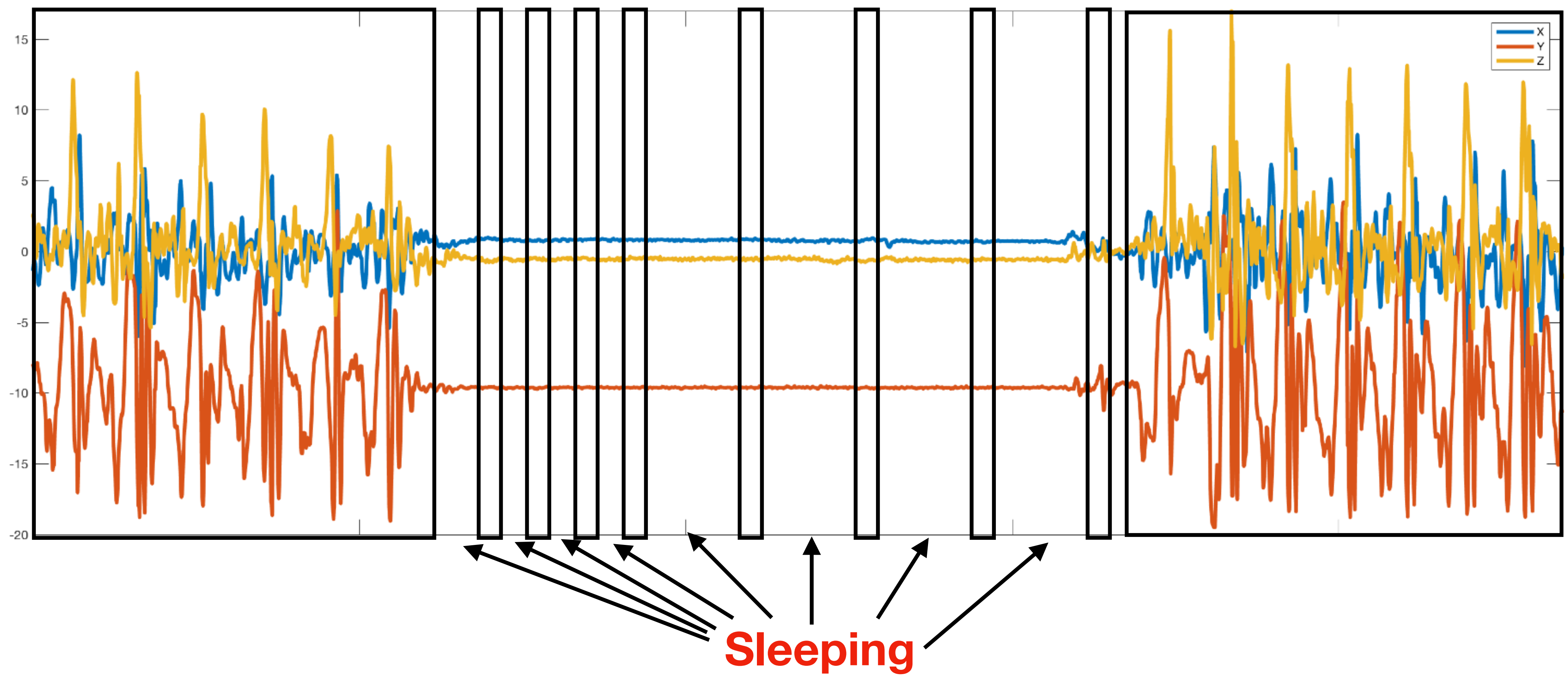
Accelerometer data

Continuous sensing challenges:

- Highly accurate data
- Very costly in terms of battery and CPU usage:
  - Continuous sensing on multiple sensors, e.g., GPS and Gyroscope, can reduce the battery life to 4-6 hours
- Can be used on cheap sensors, e.g., accelerometer

UNIVERSITY OF CAMBRIDGE

Duty Cycling

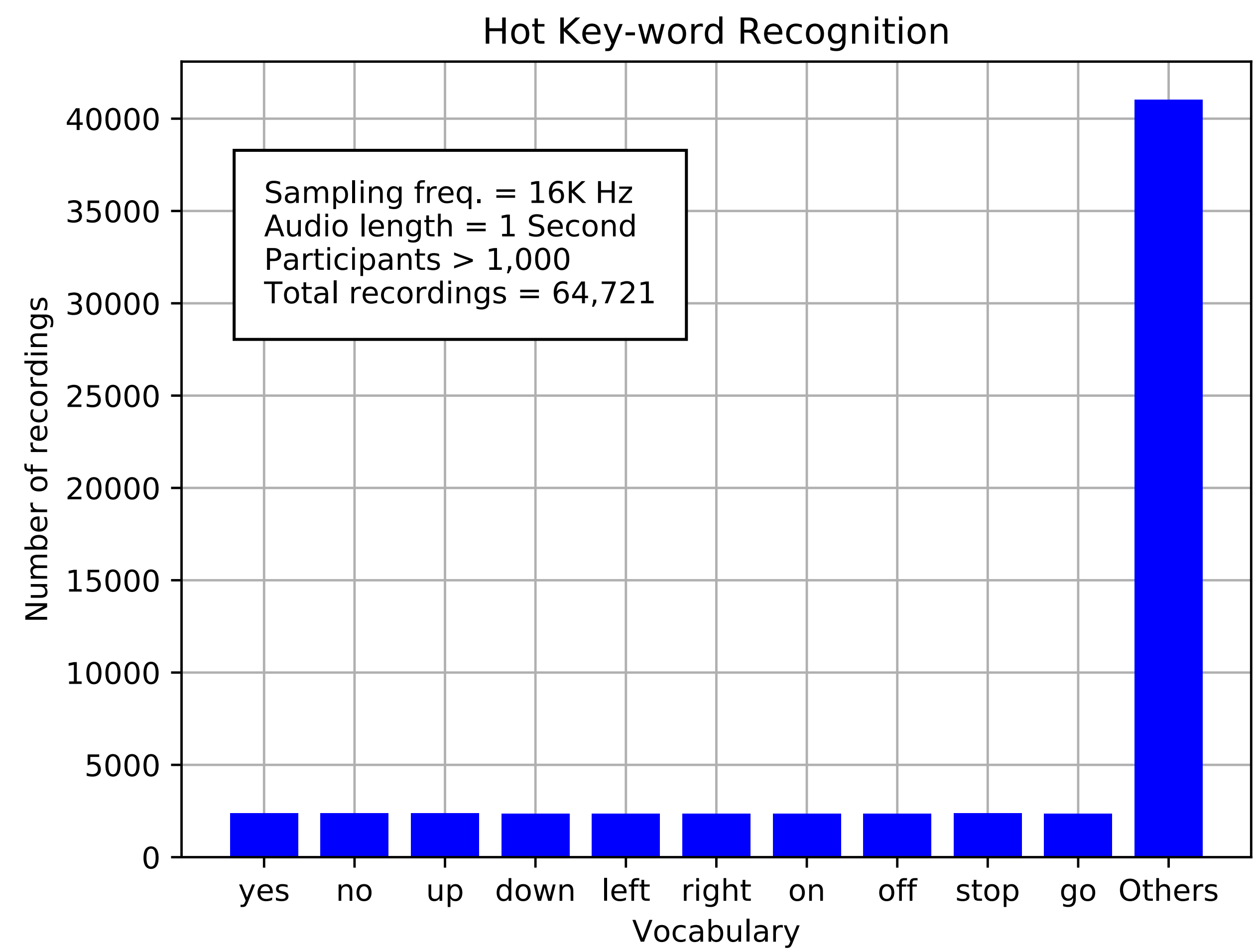


- Lower impact on the battery
- Less accurate, interesting events may take place during the sleeping period
- Adjust the duration of the sleeping periods according to the rate of the events detected
    - Sleep longer if no events are detected
    - If new events detected, reduce sleeping time

UNIVERSITY OF CAMBRIDGE

Case Study: Hot Key-word Recognition

- Recognizing a small set of 10 spoken words
- Vocabulary: yes, no, up, down, left, right, on, off, go
- Silence and Unknown
- 16 KHz, 16-bit audio



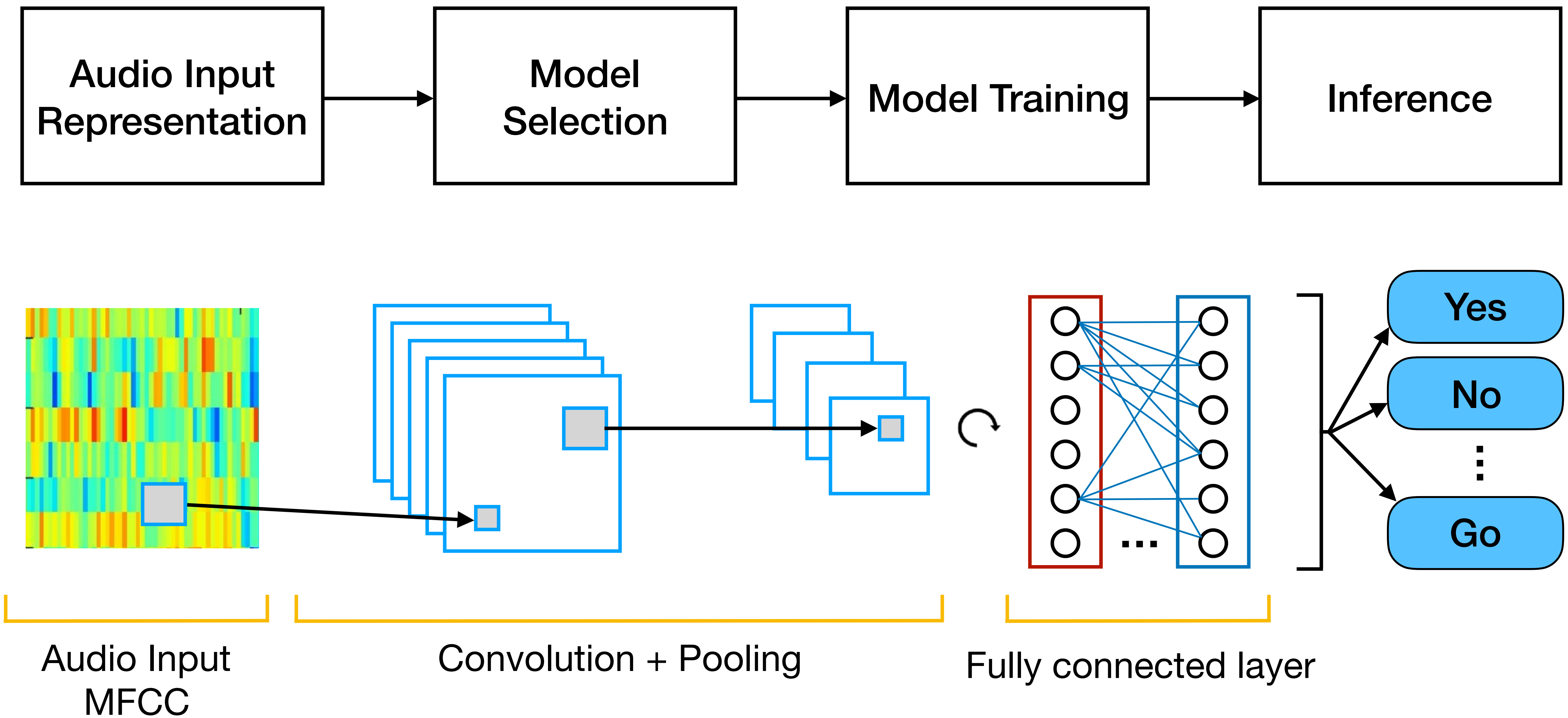
Dataset

Acknowledgement: Pete Warden, Google

https://www.tensorflow.org/versions/master/tutorials/audio_recognition

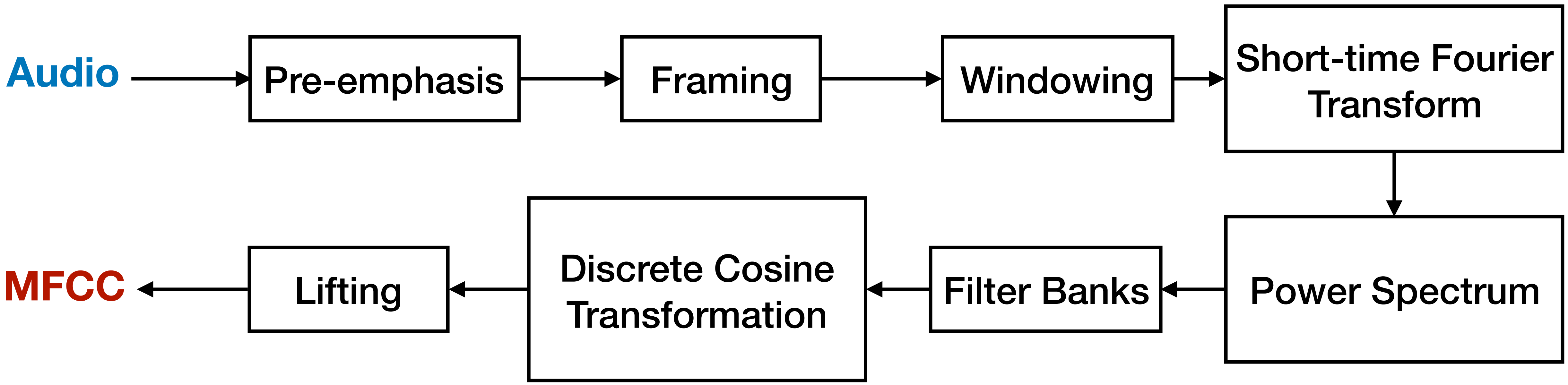
UNIVERSITY OF CAMBRIDGE

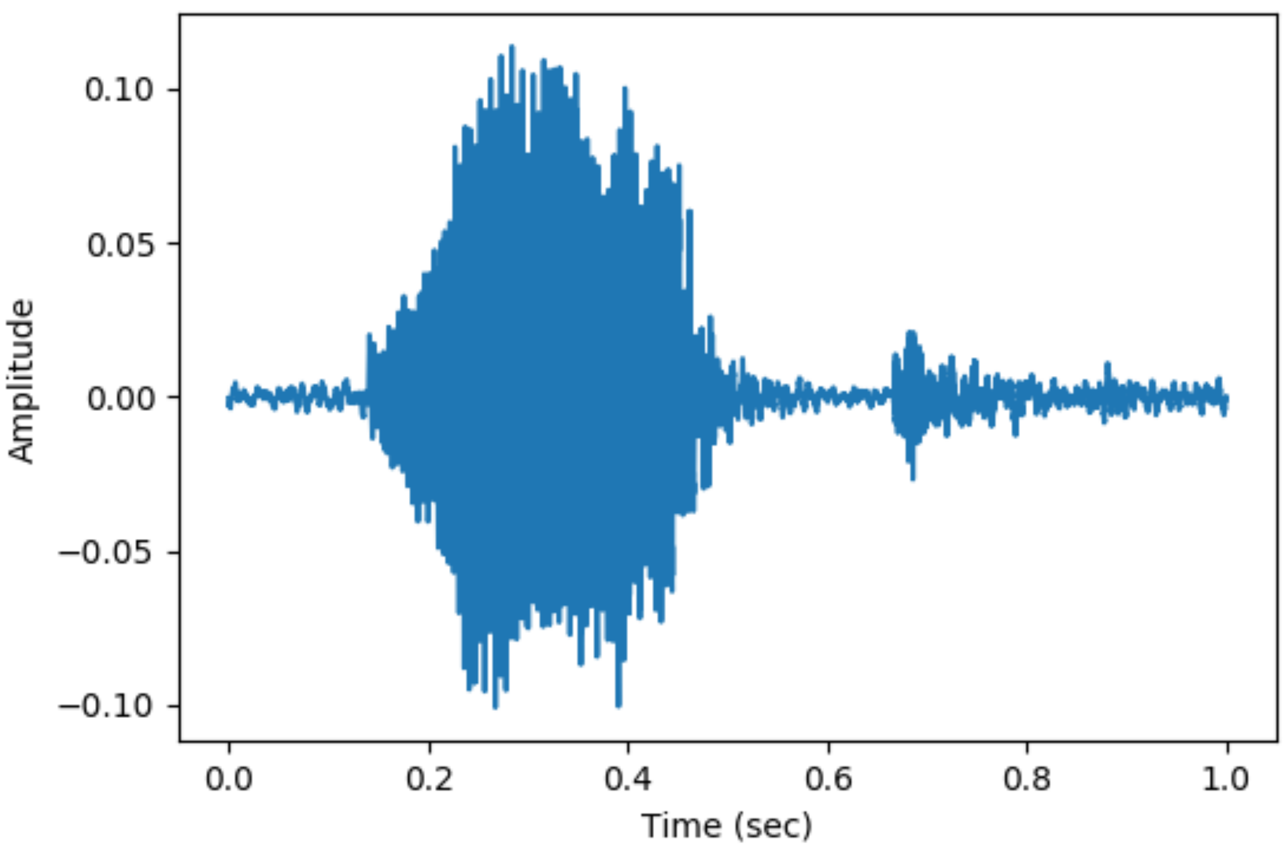
Steps in Building a Key-word Recognizer



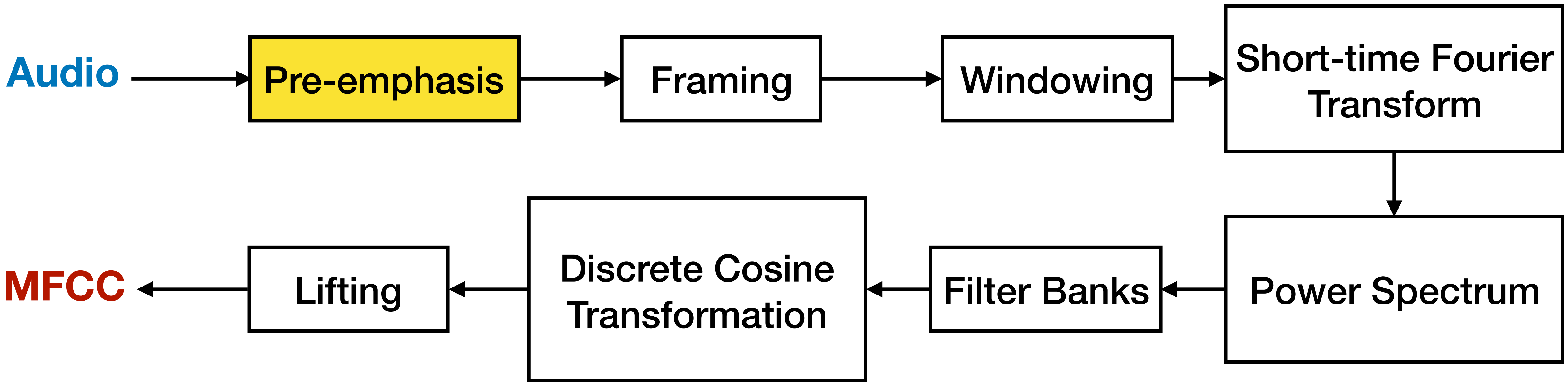
UNIVERSITY OF CAMBRIDGE

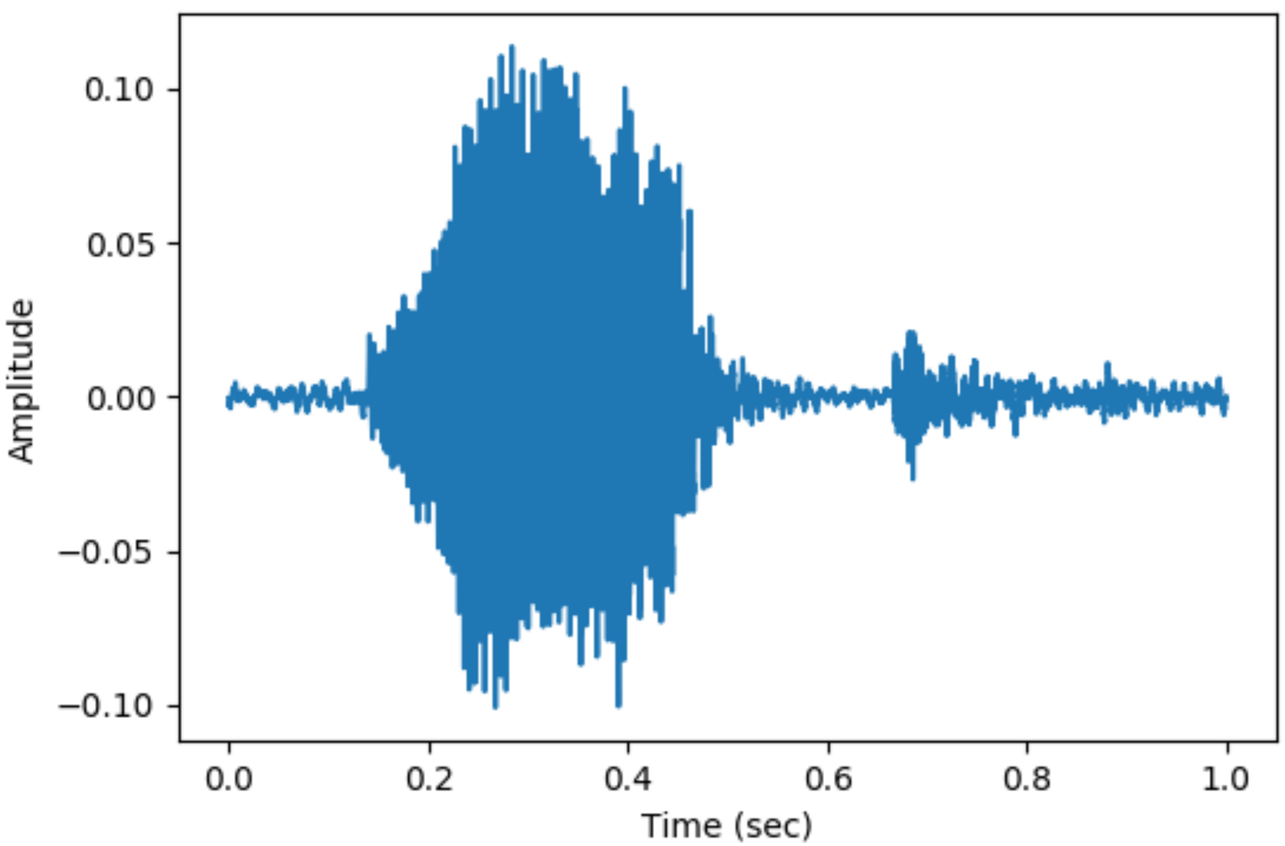
MFCC Features from Audio





MFCC Features from Audio

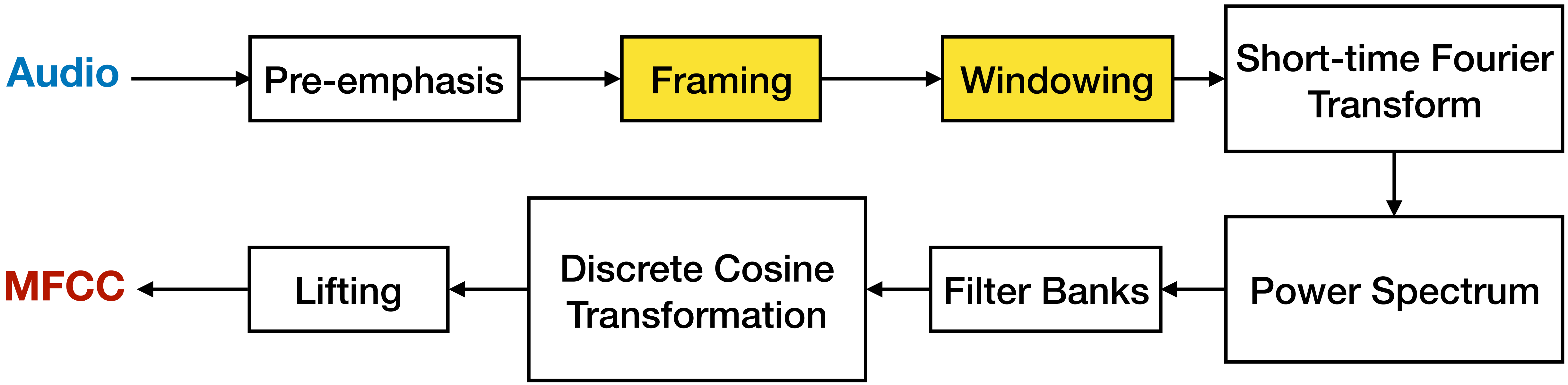


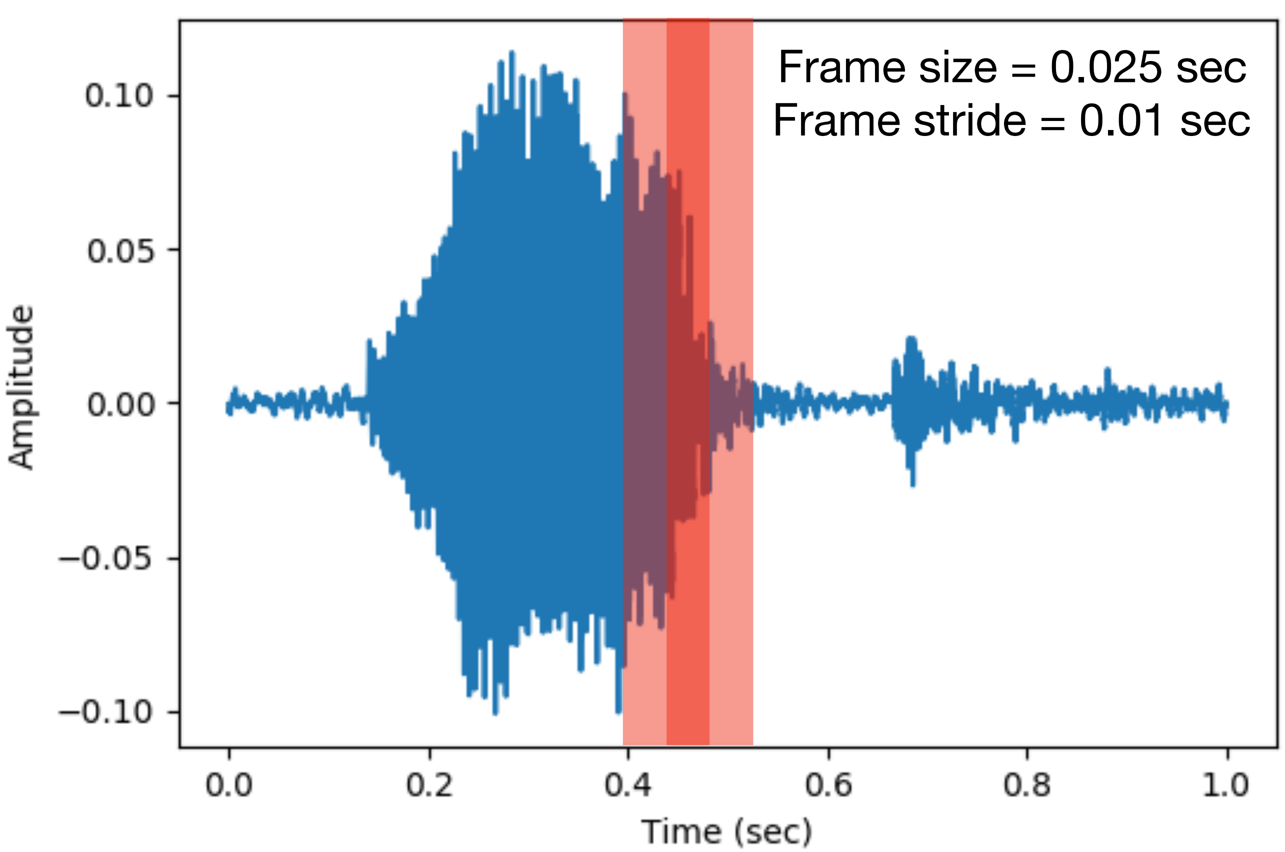


$$y(t) = x(t) - \alpha \cdot x(t-1)$$

1) Balances the frequency spectrum
2) Avoids numerical problems
3) May improve SNR

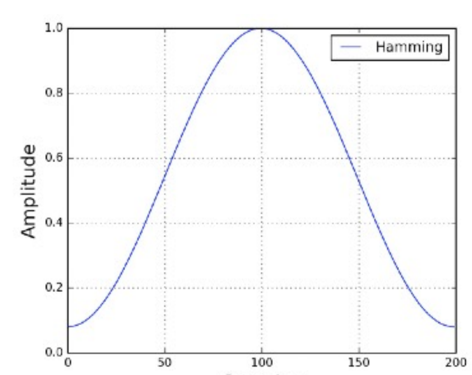
MFCC Features from Audio





Assumption:
Frequencies in a signal are stationary over a very short period of time

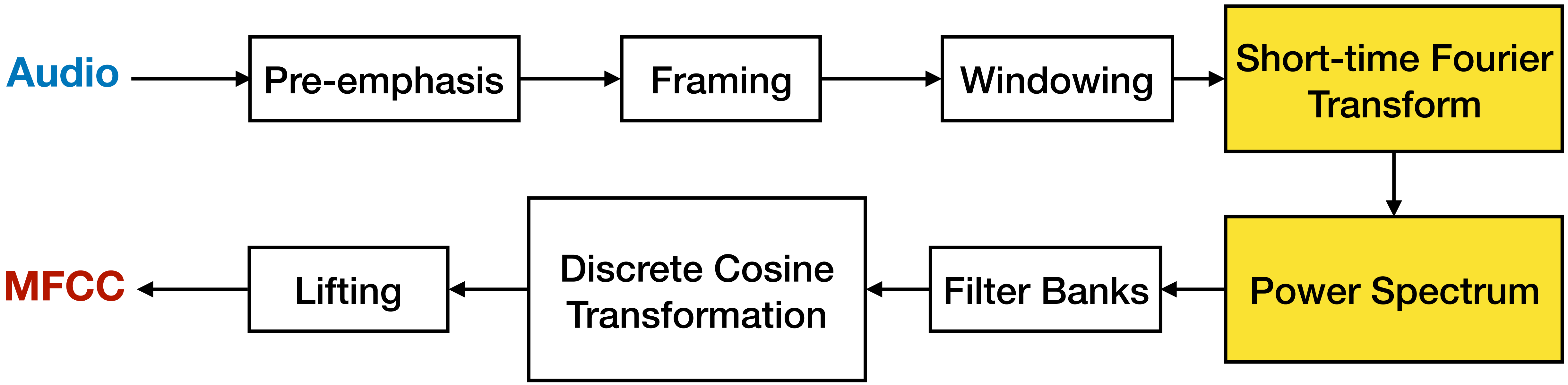
We want good approximation of the frequency contour

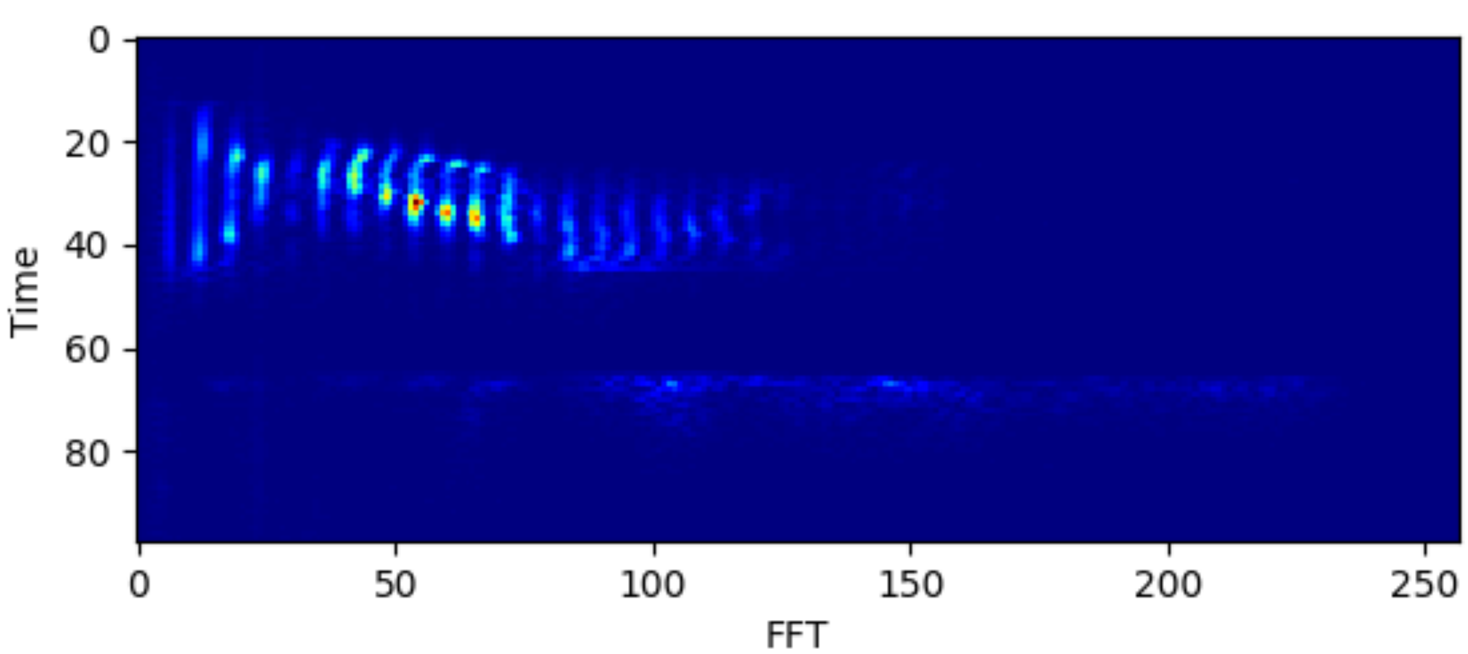


Counteract assumptions in FFT that the data is infinite

Reduce spectral leakage

MFCC Features from Audio

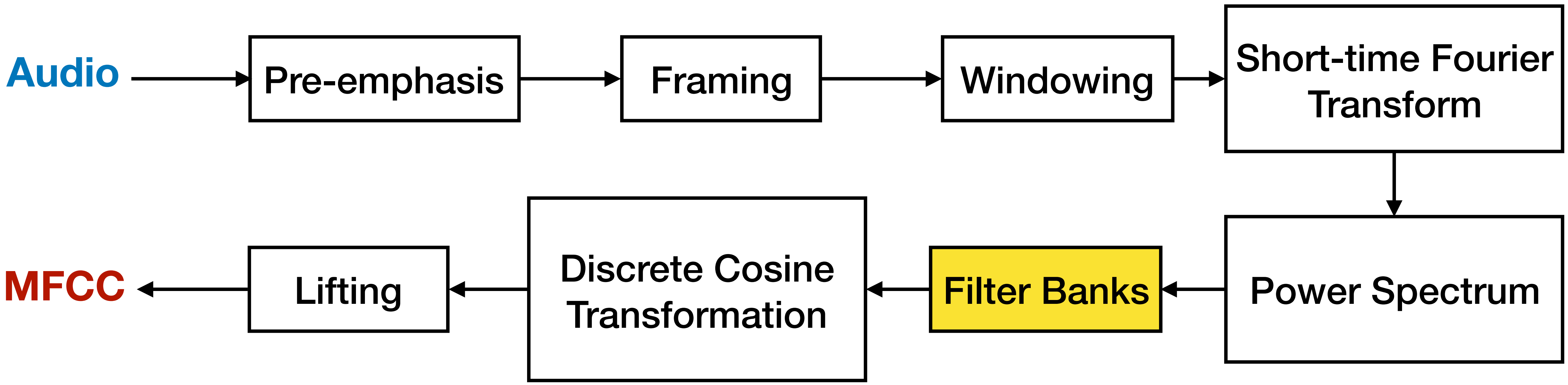


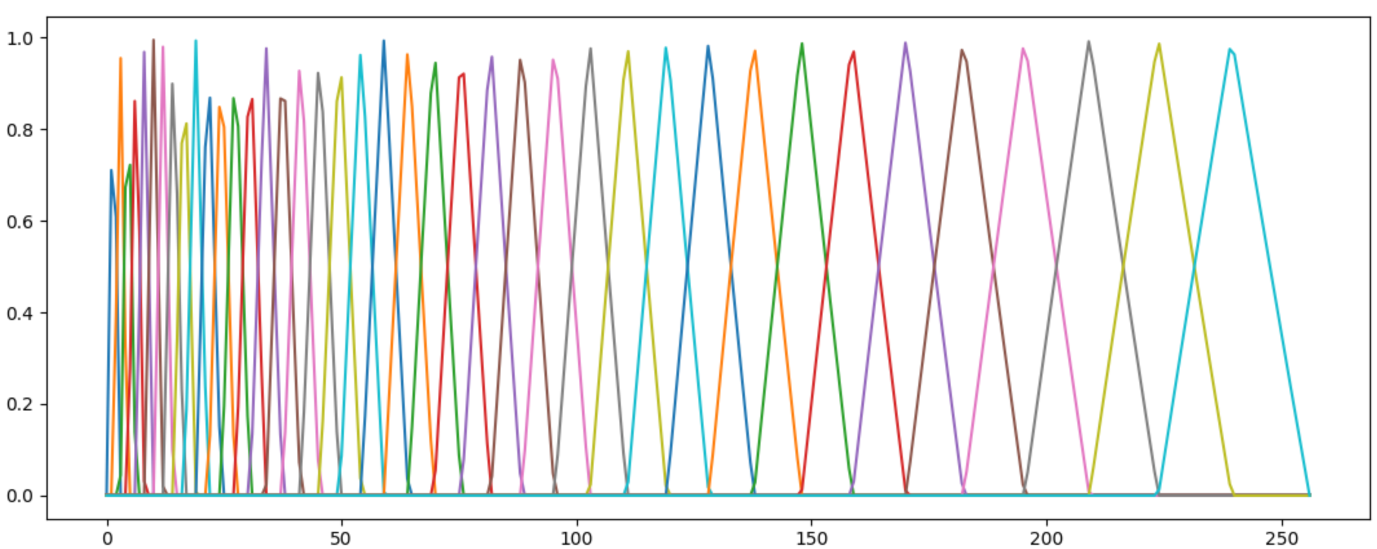


N-point FFT on each frame
N is typically 256 or power of 2

$$P = \frac{|FFT(x_i)|^2}{N}$$

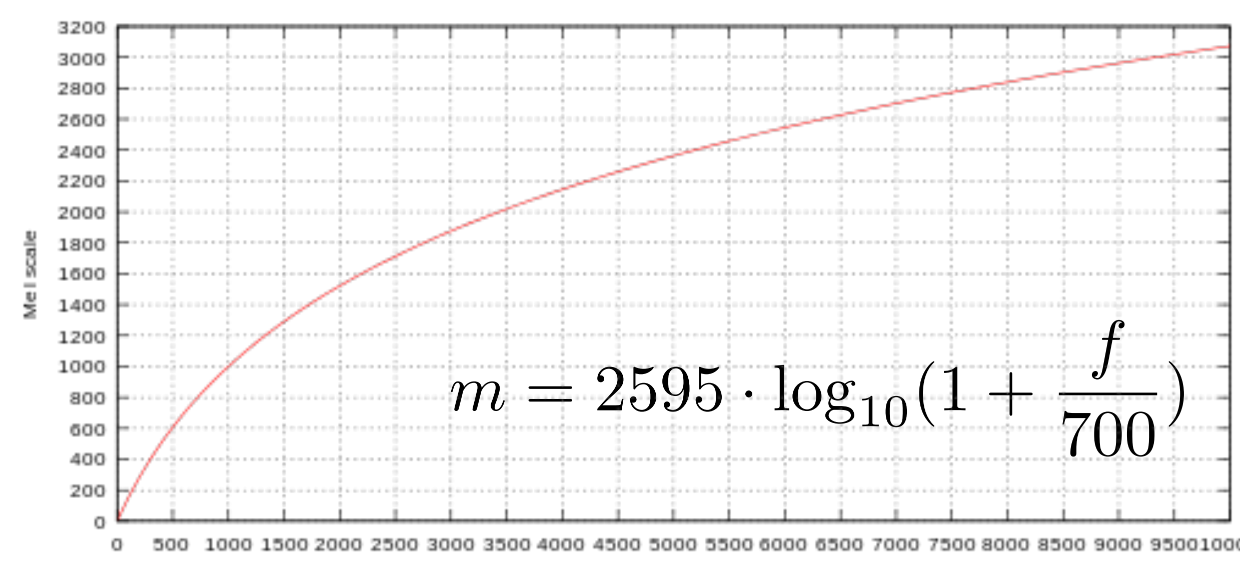
MFCC Features from Audio



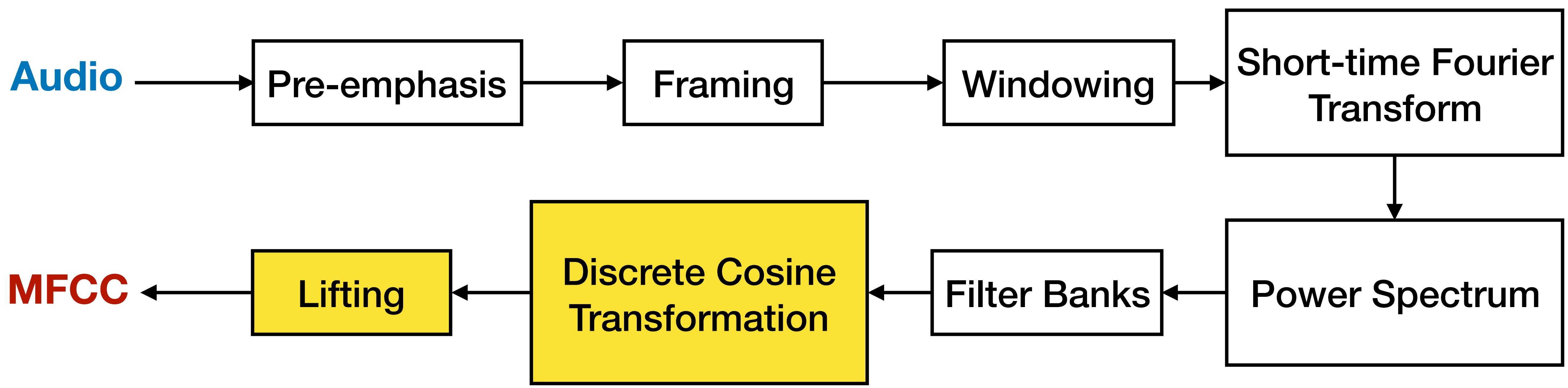


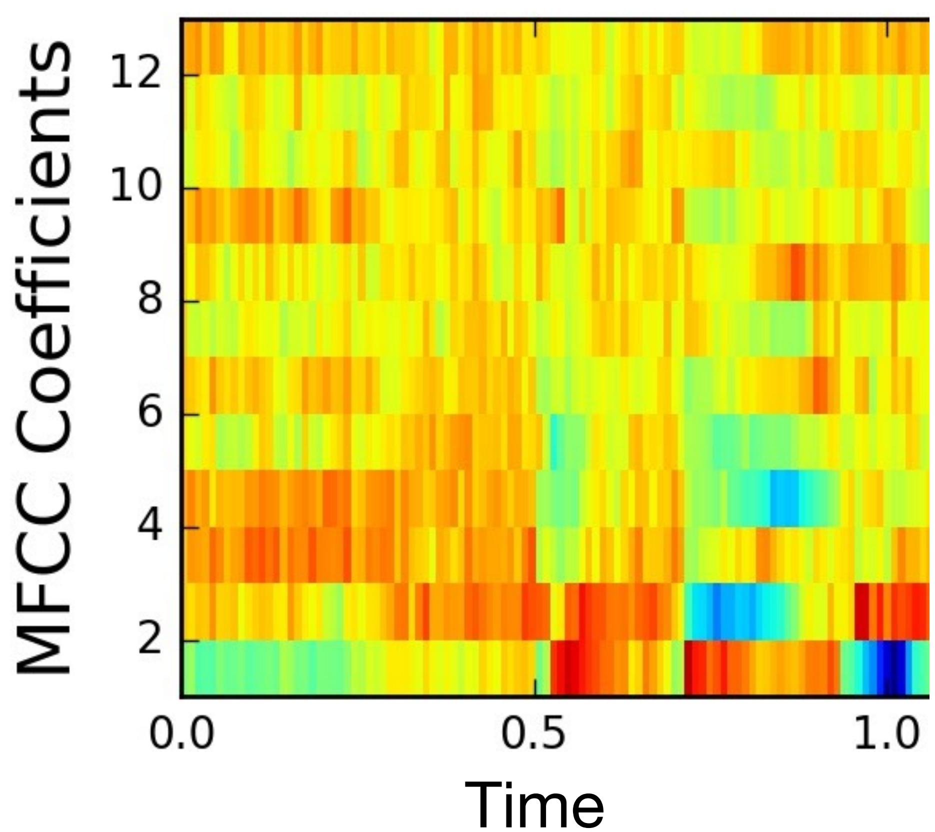
Applying triangular filters on Mel-scale

Mel-scale tries to mimic the non-linear human ear perception of sound



MFCC Features from Audio



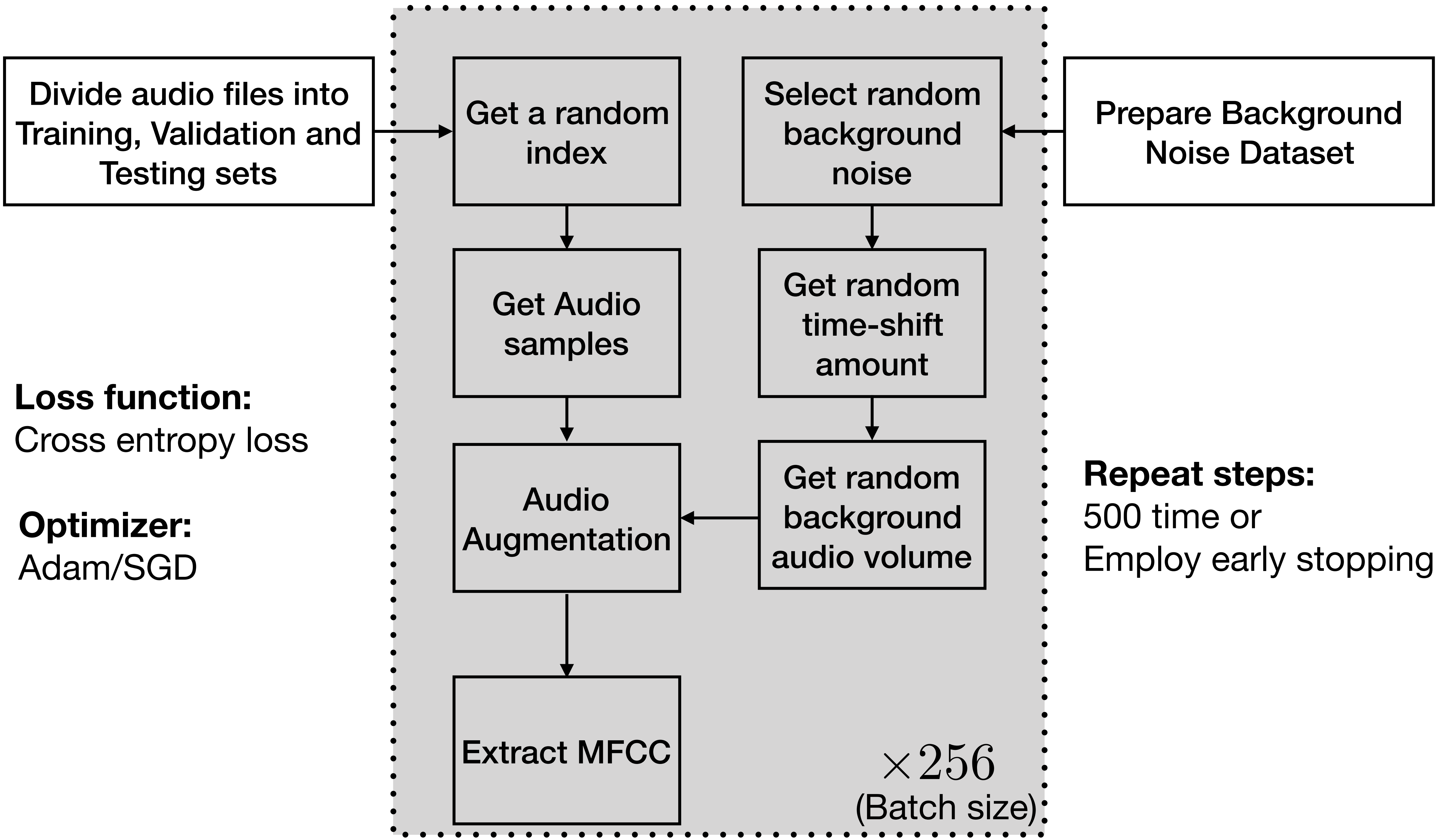


Filter-bank coefficients are highly correlated
Could be problematic in some learning algorithms

Discrete Cosine Transform (DCT) is used to decorrelate and compress filter-bank coefficients

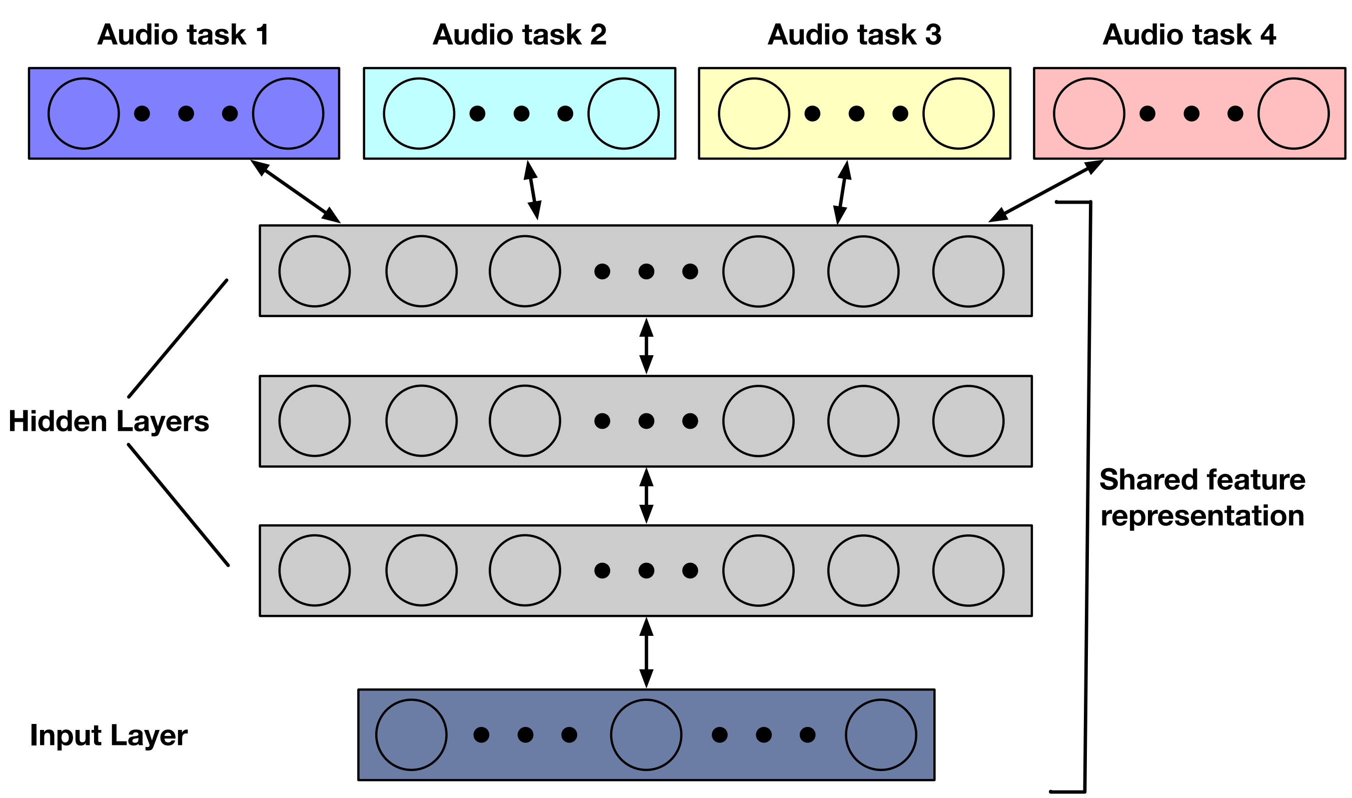
Sinusoidal lifting can be applied to de-emphasize higher MFCC coefficients

Convolutional Neural Network Training for Hot Key-word Recognition



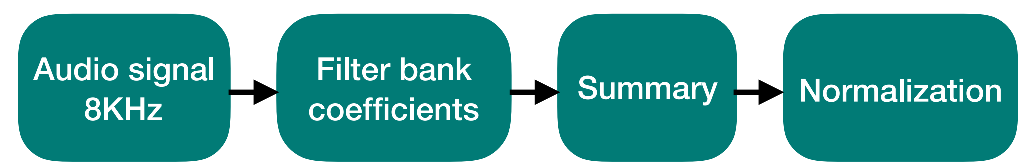
Multi-task Audio Inferencing

- **Objective:** Infer multiple contexts from the same input audio
  - Who is the speaker? Is the person stressed? Male or female speaker?
- **Multi-task learning applies an inductive transfer across domains while learning representations in parallel**

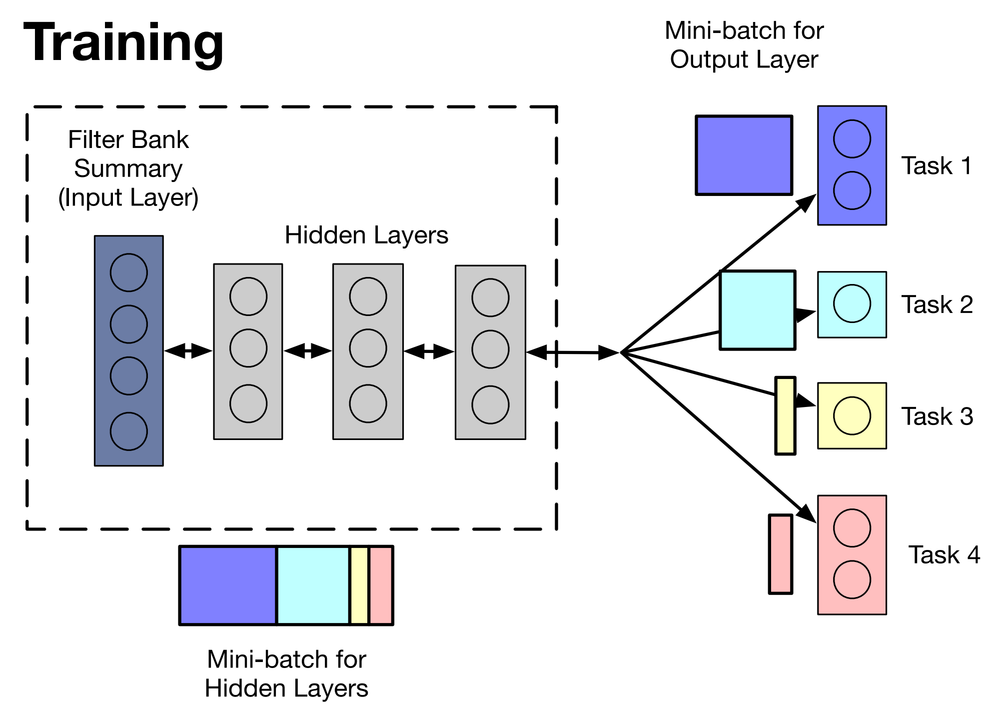


Shared hidden layer architecture

- **Audio pre-processing**



- **Training**



UNIVERSITY OF CAMBRIDGE

Open Research Questions

- How can we use unsupervised data to bootstrap the training procedure and reduce the amount of labeled data?
- How can we squeeze the resource requirements of large-scale neural networks for resource-constrained devices?
- Protecting privacy of the users.
- Multi-modal rich modeling of sensor data for accurate high-level context-recognition.

UNIVERSITY OF CAMBRIDGE

References

- N.D. Lane, E. Miluzzo, H. Lu, D. Peebles, T. Choudhury, A. Campbell. A survey of mobile phone sensing. IEEE Computer Magazine. Vol. 48. No 9. September 2010.

- K.K. Rachuri, M. Musolesi, C. Mascolo, P.J. Rentfrow, C. Longworth, A. Aucinas. EmotionSense: A Mobile Phones based Adaptive Platform for Experimental Social Psychology Research. Ubicomp'10. September 2010.

- S. Hemminki, P. Nurmi, S. Tarkoma, Accelerometer-based transportation mode detection on smartphones, SenSys 2011.

- S. Bhattacharya and N.D. Lane, Sparsification and Separation of Deep Learning Layers for Constrained Resource Inference on Wearables, SenSys 2016.

- P. Georgiev, S. Bhattacharya, N.D. Lane, C. Mascolo, Low-resource Multi-task Audio Sensing for Mobile and Embedded Devices via Shared Deep Neural Network Representations. IMWUT (UbiComp) September 2017.

UNIVERSITY OF CAMBRIDGE