Task Introduction

Task: Multiclass Classification

Framewise phoneme prediction from speech.

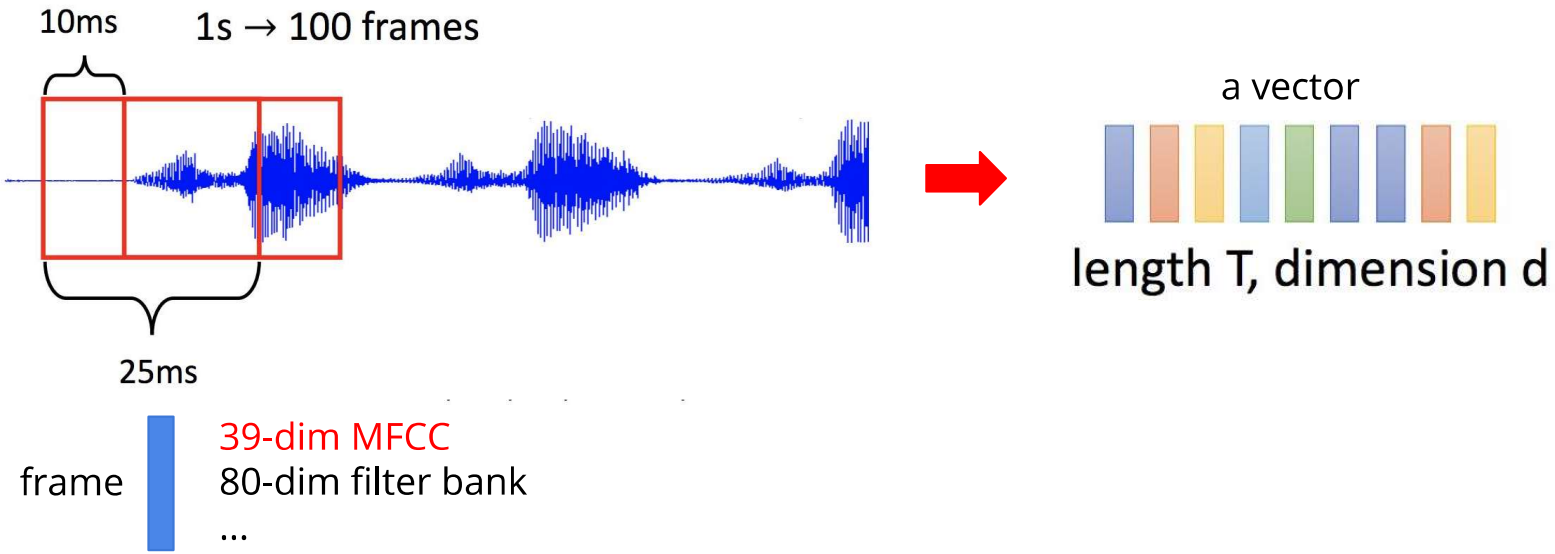
What is a phoneme?

A unit of speech sound in a language that can serve to distinguish one word from the other.

- bat / pat , bad / bed
- Machine Learning → M AH SH IH N L ER N IH NG

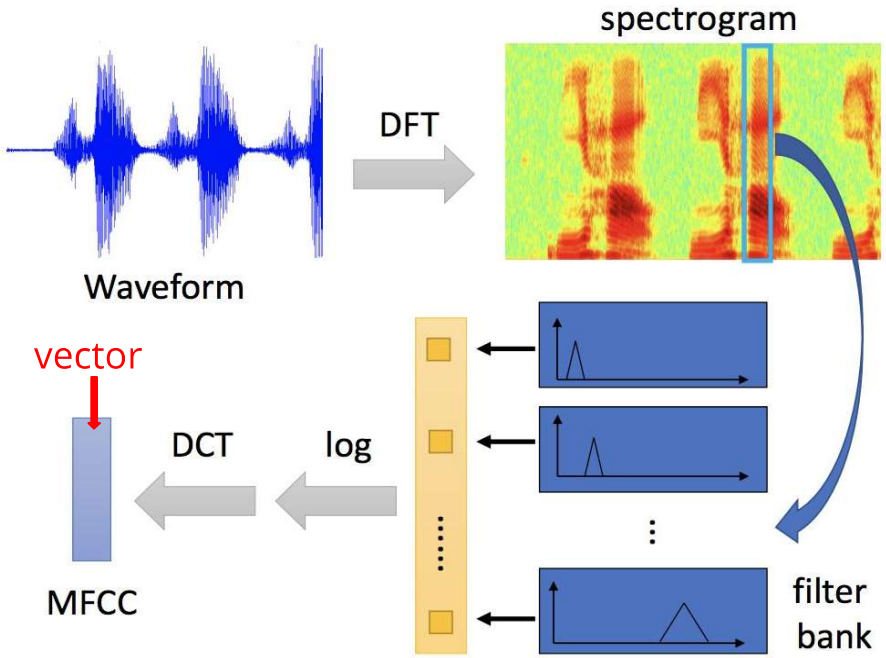
Task Introduction

Data Preprocessing



Task Introduction

Acoustic Features - MFCCs (Mel Frequency Cepstral Coefficients)



For more details,
please refer to Prof. Lin-Shan Lee's
[Introduction to Digital Speech Processing] Chap.7

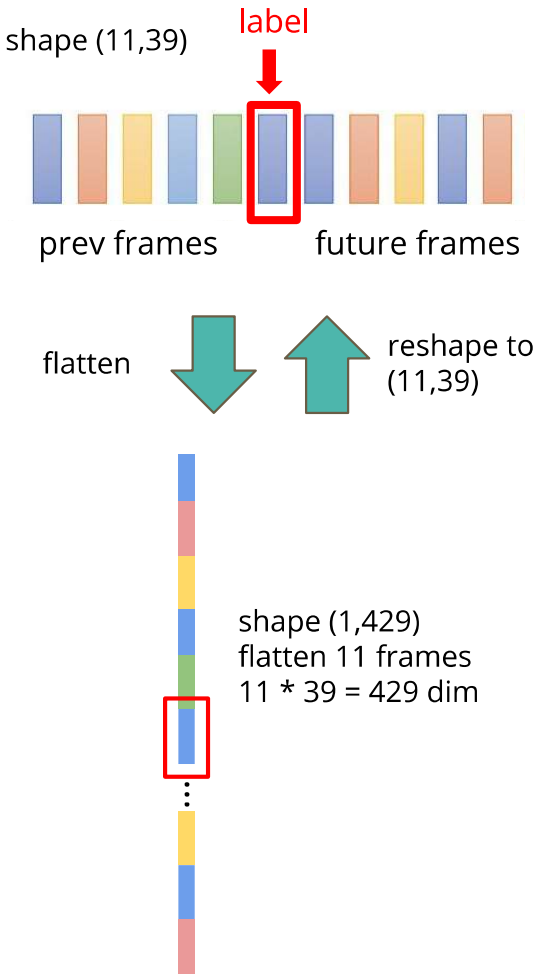
Image ref.
Prof. Hung-Yi Lee
[2020Spring DLHLP] Speech Recognition

More Information About the Data

Since each frame only contains 25 ms of speech, a single frame is unlikely to represent a complete phoneme

- Usually, a phoneme will span several frames
  - Hint: post-processing may help
- Concatenate the neighboring phonemes for training
  - In this HW, we concatenate the past and the future five frames for training (total 11 frames)
  - You may reshape the input (1,429) back to (11,39) to get separated 11 frames
  - Just remember that the label corresponds to the center frame
- Finding testing labels or doing human labeling are strictly prohibited!

Introduction to Digital Speech Processing



Dataset & Data Format

- Dataset: TIMIT Acoustic-Phonetic Continuous Speech Corpus
  - Phonetically balanced for English
- Data Format (The TAs have already preprocessed the data)

  timit_11/
  - **train_11.npy → training data (# of training frames, 11 x feature dim)**
  - **train_label_11.npy → framewise phoneme label (0-38)**
  - **test_11.npy → testing data (# of testing frames, 11 x feature dim)**
  - Acoustic features (39-dim MFCC)
    - Concatenate the past and the future five frames (feature dim = 11 x 39)
    - The phoneme label of each input corresponds to the center frame
- **Using additional data is prohibited**. Your final grade will be multiplied by 0.9!

Class	Phoneme	Example	Class	Phoneme	Example	Class	Phoneme	Example
0	iy	b***ee***t	13	l	***l***ay	26	dx	mu***dd***y
1	ih	b***i***t	14	r	***r***ay	27	g	***g***ay
2	eh	b***e***t	15	y	***y***acht	28	p	***p***ea
3	ae	b***a***t	16	w	***w***ay	29	t	***t***ea
4	ah	b***u***t	17	er	b***ir***d	30	k	***k***ey
5	uw	b***oo***t	18	m	***m***om	31	z	***z***one
6	uh	b***oo***k	19	n	***n***oon	32	v	***v***an
7	aa	b***o***b	20	ng	si***ng***	33	f	***f***in
8	ey	b***ai***t	21	ch	***ch***oke	34	th	***th***in
9	ay	b***i***te	22	jh	***j***oke	35	s	***s***ea
10	oy	b***o***y	23	dh	t***h***en	36	sh	***sh***e
11	aw	b***ou***t	24	b	***b***ee	37	hh	***h***ay
12	ow	b***oa***t	25	d	***d***ay	38	sil	silence/closure sounds