Exercise 3: Voice Activity Detection

Mohammad Vali

Aalto University, Finland

DeadLine: Monday, September 25, 2023

Instructions

- Implement and return files as **Exercise_3_firstname.ipynb**.
- Return your answers to MyCourses by **23:59 on Monday, September 25th**, 2023.

Introduction and Objective

- Learn and understand concepts about voice activity detection (VAD) practically.
- VAD: Detect the presence or absence of speech in an audio sample.
- Application: Almost all the speech technologies !!
- You will be implementing three functions. They are,
  (1) zcr –> Function for zero crossing rate.
  (2) energy –> Function for energy.
  (3) add_deltas_deltadeltas –> Function for deltas and double deltas.
- Some functions are already provided in the note book to reduce the work load. They are: F0_autocorr, F0_cepstrum, read_ground_truth, normalize and perceptron_training

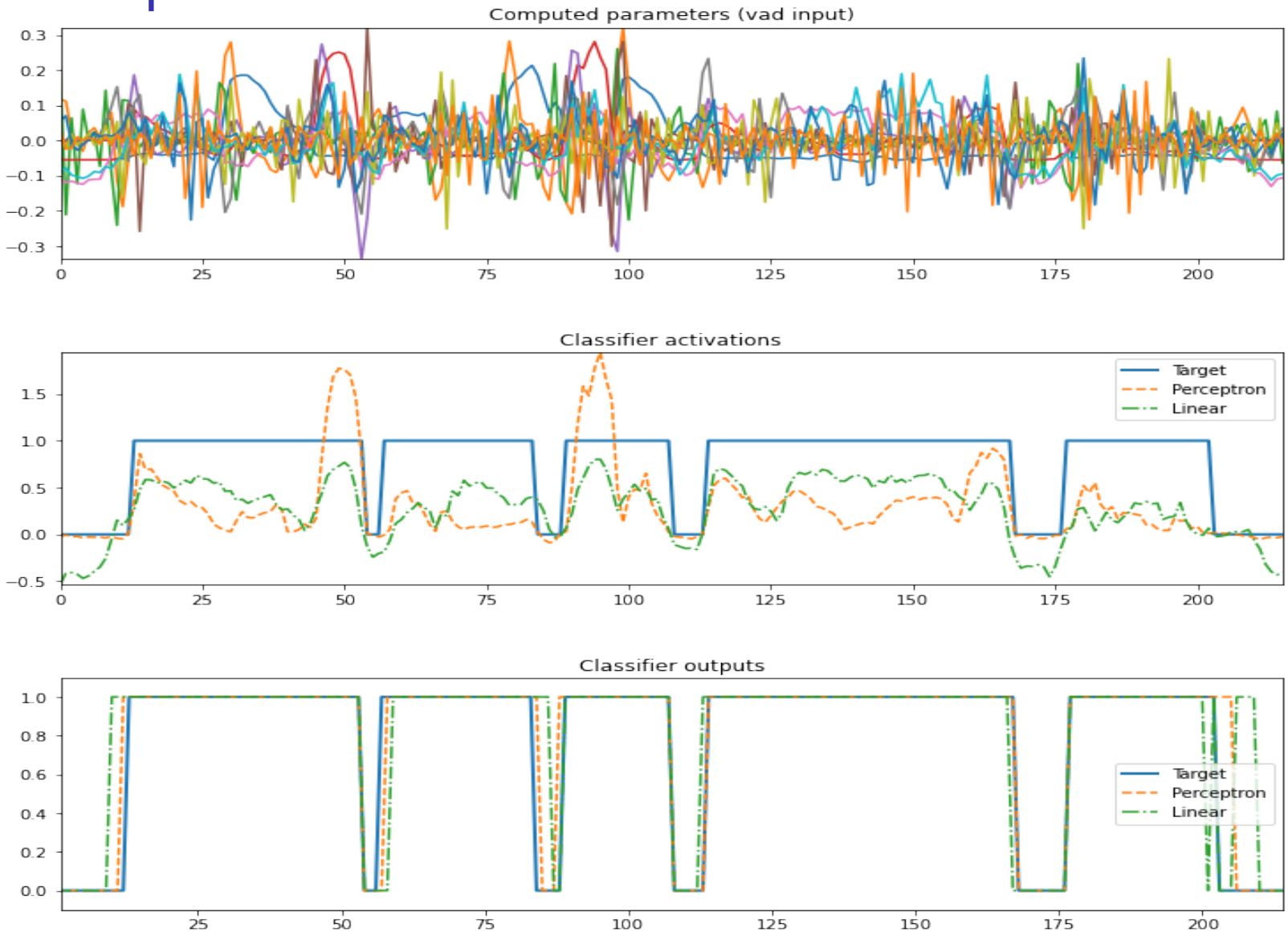
Features and Classifiers

- Features:
  - Zero Crossing Rate (ZCR)
  - Frame Energy
  - F0 frame peaks using autocorrelation (previous exercise).
  - F0 frame peaks using cepstrum (previous exercise).
  - Delta and delta-delta features.
- Classifiers: Linear classifier.

Procedure

- Compute the features
- Train the classifiers
- Use trained classifiers for classification.
- Evaluate: Mean Absolute Error
- Plot

Expected plot



Experiment with the parameters

- What is the single best parameter for VAD ?
- What is the worst performing parameter?
- What other features could be useful for VAD?
- What do you think the challenging aspect in VAD task?

Test your VAD detector on your own voice (Optional)

- Record your own speech file, then create a ground truth and then run this notebook for your speech file. Write your observations.

Learnings

Experimental findings, Analysis, Reasoning and and any other?

Contact

- **Email:** mohammad.vali@aalto.fi