NI-MVI Semestra project 2022

Goal:

Compare the quality of audio recording of Spanish speakers enhanced by CMGAN model. I will compare inferred data from :

- A. Model pretrained on english speakers as provided by authors of the CMGAN.
- B. Fine-tuned pretrained model with Spanish speakers.

Model:

CMGAN: Conformer-Based Metric-GAN for Monaural Speech Enhancement Sherif Abdulatif, Ruizhe Cao, Bin Yang

- Paper: ArXiv

- Implementation: GitHub

- My testing environment: Google Colab

Framework: PyTorch

Data:

Spanish speaking audio recording in high quality (podcast quality).

Data downloaded from YouTube.

Both sexes - male and female.

After preprocessing 50 minutes of data.

Train:evaluation ratio is 40:10.

Preprocessing consists in:

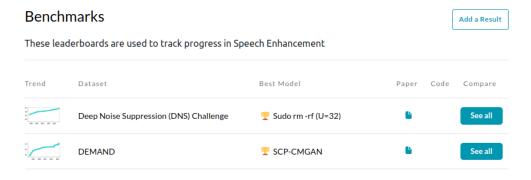
- Tokenization: split data to short audio files, 3-8 seconds long.
- Downsampling: recommended procedure in paper, 16kHz and 16 bits per sample.
- Adding noise using the <u>DEMAND dataset</u> as recommended in the paper.

List of data sources is in the *sources.txt* file.

The preprocessed data are available on my university <u>Google Drive</u>, link in *sources.txt*. Preprocessing notebook: <u>GitLab</u>

Research:

CMGAN is almost SoA. The successor SCP-CMGAN offers other metrics system which I did not understand so I chose the closest solution with available pretrained model and public dataset. paperswithcode.com



Approach:

- 1. Infer evaluation data with the available pretrained model
 - a. Compute score (PESQ and STOI).
- 2. Fine-tune existing pretrained model using custom data with Spanish speakers.
 - a. Infer evaluation data.
 - b. Compute score.
- 3. Compare results.