BERNAS Raphaël
CORLAY Maxime
LETELLIER Adrien
ZHENG Emilie

Project Task Repartition:

NLP

Subject : Augmentation Invariant Discrete Representation for Generative Spoken Language Modeling.

Motivation : A Speech-to-Speech system is often divided into three main components : Speech-to-Unit, Unit Language Model (LM), and Unit-to-Speech. In the paper "Augmentation Invariant Discrete Representation for Generative Spoken Language Modeling", the author primarily focuses on the Speech-to-Unit component. This component consists of an encoder and a quantizer model, which take a traditional continuous sound signal as input and produce a discrete output. Several methods have been proposed for generating these discrete outputs, but most have shown robustness issues. To address this, the authors propose a processing method for the quantizer to improve the quality of the outputs. They introduce a metric called Unit Edit Distance (UED), based on the Levenshtein distance, which compares two signals processed into discrete outputs and returns a similarity score. Using this metric, they define a loss function to compare a k-means quantizer with a Multi-Layer Perceptron (MLP) quantizer, updating the MLP quantizer to improve its performance based on this metric. This process is repeated iteratively to develop progressively better quantizers. Unfortunately, this approach has limitations. In their paper, the authors discuss the architectural choices for the initial quantizer, but the validity of the metric itself is also questionable, particularly in a Speech-to-Speech context, where measuring word distance accurately is challenging.

Task repartition:

a) To recode the encoder part to have the basics [Adrien]

b) To recode the quantizer part with the training method introduced in their paper [Raphaël]

c) To recode testing methods: UED and ABX [Raphaël]

d) To implement general augmentation and some more [Maxime]

e) To implement new augmentation (corruption and volume change) [Emilie]

f) To test models on the original settings [Emilie]

g) To test new models [Maxime]

h) To suggest other metric, mainly reform the UED metric and Levenshtein distance using Dynamic Time Warping [Adrien]

i) To test new architecture of quantizer for E_1 (change the MLP based method) [Raphael]