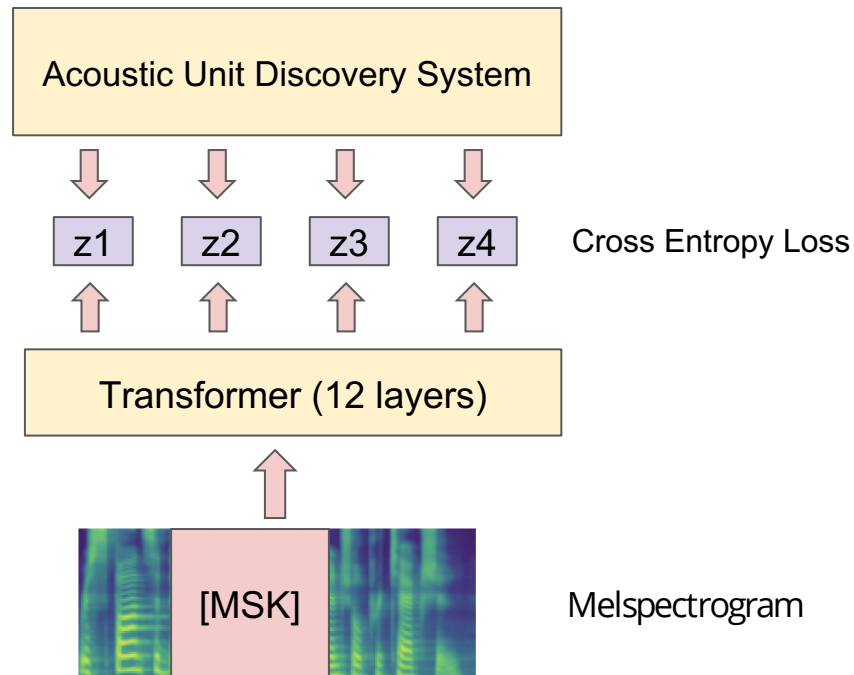# JSALT Team Meeting

Tzu-Quan Lin
2022/06/12

We train our own Mel-HuBERT

- Mel-HuBERT is quite similar to the original HuBERT, but we **remove the convolutional feature extractor** to speed up the whole process
- The convolutional feature extractor actually account for about **33%** computational overhead in original HuBERT, so we believe that removing the CNN extractor will help us accelerate the model

Acoustic Unit Discovery System

z1   z2   z3   z4   Cross Entropy Loss

Transformer (12 layers)

[MSK]   Melspectrogram

Why don't we use the pretrained model of fairseq?

- We think there are pros and cons using pretrained model of fairseq
  - Pros
    - Much more easier, do not need too much computational resources to train from scratch
    - Our baseline model would be HuBERT Base, which is one of the SOTA speech ssl model
  - Cons
    - We lose the transparency and control over the pre-training procedure of our baseline model
    - We are not 100% sure whether every implementation details are written as we expect

Stage 1 Pre-training Result on Libri-360

	Phone Recognition (PER%)	Speaker Identification (ACC%)
50 epochs one	23.04	63.59
150 epochs one	16.08	65.81
200 epochs one	15.10	64.75

Stage 1 pre-training experiments on Libri-360

Stage 2 pre-training experiment

	Original	10 epochs one	50 epochs one	100 epochs one
Phone Recognition(PER%)	17.25	14.50	13.62	12.35

Stage 2 pre-training experiments on Libri-360

Preliminary Result on Libri-960

	Phone Recognition (PER%)	Speaker Identification (ACC%)
Libri-360, 200 epochs	15.10	64.75
Libri-960, 100 epochs	13.68	72.48

Stage 1 pre-training on Libri-360 and Libri-960

I am currently working on…

- Waiting for the result of stage 1 and stage 2 pre-training on Libri-960, we plan to scale up every experiments to Libri-960 and stage 2 pre-training model for about 1 month later

Model checkpoint releasing timeline

- Model checkpoints will be released at this link
  - Already available: Mel-HuBERT stage 1 pre-training checkpoint on Libri-360
  - 6/20: Mel-HuBERT stage 1 pre-training checkpoint on Libri-960
  - 7/11: Mel-HuBERT stage 2 pre-training checkpoint on Libri-960
- See this README.md to know more about how to use Mel-HuBERT

Thanks for listening!