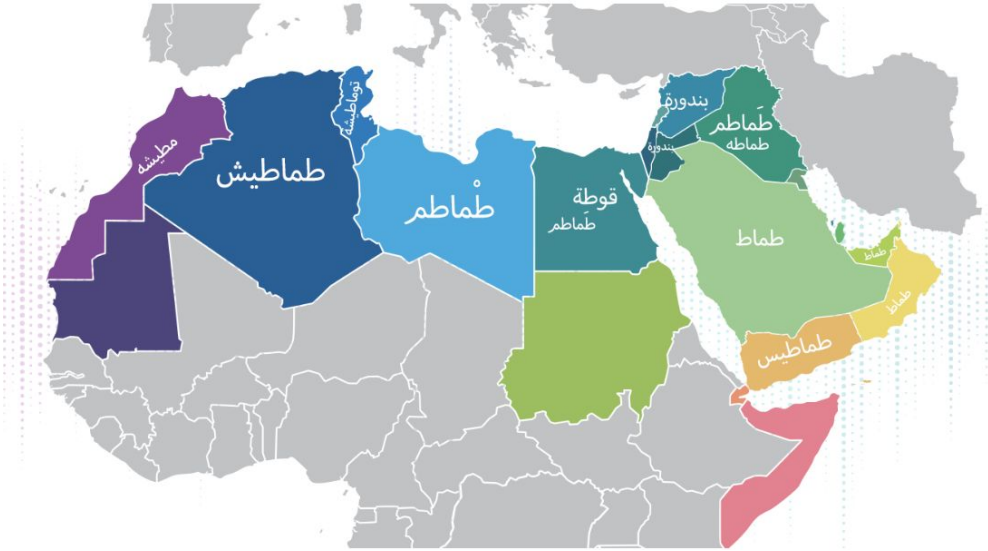
Arabic Dialect Identification / Diarization using Wav2Vec 2.0



Daniel Izham, Phoebe Sun, Yassine EL Kheir and Youssef ELShahawy

Mentor: Markus Muller

Overview I - Arabic ASR

- Arabic
  - spoken by 300 million people
  - 17 dialects
  - Similar vocabulary and grammar
  - Different pronunciation, colloquial expressions, and cultural references
- Multi-dialect arabic ASR is a hard task!

Overview II - Arabic Dialect Identification / Diarization

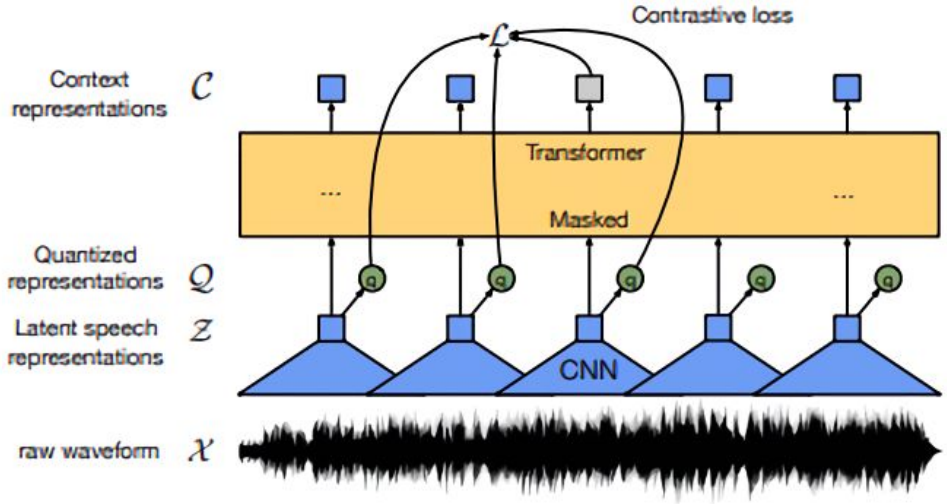
- Dialect identification / diarization
  a. Improved accuracy
     - tailor its transcription to the specific dialect being spoken
  b. Enhanced robustness
     - better handle variations in pronunciation, vocabulary, and grammar that may be present in different dialects
  c. Increased efficiency
     - use more specialized models or processing techniques that are better suited to the specific dialect

Social and economic impacts

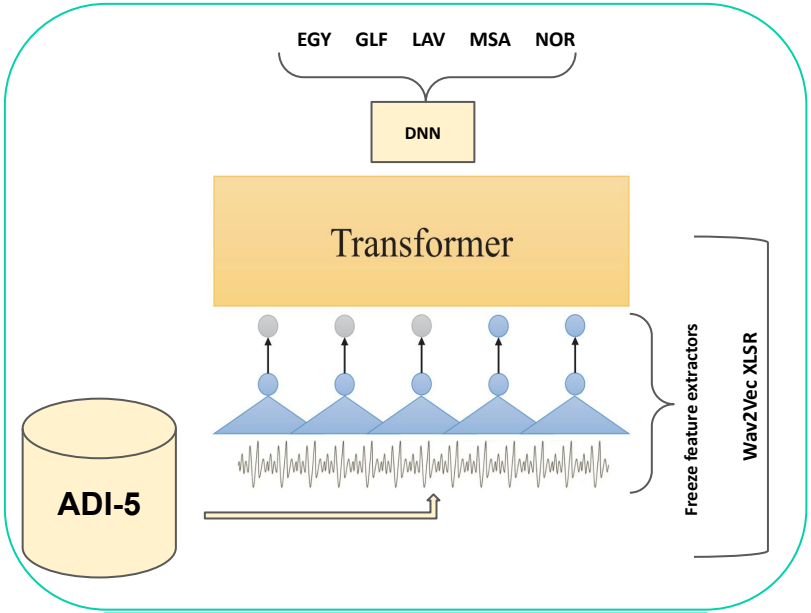
- Facilitate communication and understanding among people from different Arab regions who speak different dialects
- Facilitate trade, education, and other forms of social and economic exchange
- Foster greater cultural awareness and appreciation, and can help to build stronger social and economic ties between different Arab regions and communities
- Help to preserve linguistic diversity within the Arab world

Why fine-tuning Wav2Vec2.0 for arabic dialect identification?

- Experiment with the Wav2Vec 2.0 compared to i-vectors, MFCC, F0, FBANK, Spectrograms which produced state of the art results.
- Wav2Vec is able to process raw audio waveforms and provide features automatically
- Robustness of self-supervised model.



Data and Model Fine-tuning Overview

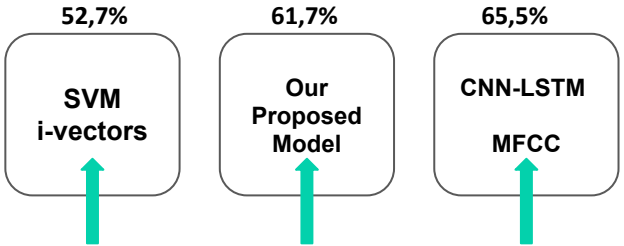


Model Architecture

Overfitting:

- Training with different chunk duration of audios from 1s-9s
- Perturbation of the original dataset in term of speed slightly faster or slower

Accuracy and Comparison:



Challenges/Results

Results

1. Jupyter notebook tutorial
2. Possible adjustments for offline use
3. Streaming website / demo
4. The model will be uploaded to the Hugging Face
5. Use this tuned model’s weights to train multi-dialect arabic ASR

Demo

https://www.youtube.com/watch?v=RmjBfgCU0II