Improving Mispronunciation Detection and Diagnosis for Non-native Learners of the Arabic Language

Nina Hosseini-Kivanani, Norah Alrashoudi, & Aswin Giridhar Subramanian

Mentor: Stan Salvador

08/01/2023-SLT2022-Doha

Goals

- Pronunciation plays an important role in L2 learning process,
  - due to a lack of time and resources compared to other skills in classrooms
- To support learning of Arabic language and help non-native Arabic speakers to improve their pronunciation skills.
- Evaluating the efficiency of Wav2Vec model for training Arabic dataset

Social Impacts

- Help language teachers and improve autonomous learning for language learners in their pronunciation progress
  - Improving the e-learning experience

Dataset Description

Information	Dataset	Used part
Language	Arabic	Arabic
Speakers	40 Native Arabic & 40 African L1 (learn Arabic)	10 African L1 (learn Arabic)
Level	Intermediate or advanced learners	Intermediate learner
Data	Read sentences	Read sentences
Hours	4,000 utterances, 5 hours and 33 minutes	150 utterances, 19 minutes
Age	Adult	Adult
Annotation	Word & phoneme level	Word & phoneme level
Mispronunciation	Deletion/addition/substitution	Deletion/addition/substitution
Labeling	Al-Tamimi Romanization of Arabic Orthography (ATR)	Al-Tamimi Romanization of Arabic Orthography (ATR)

Dataset Description

Arabic	English Transliteration
الحمدُ للهِ ربِ العالمين الرحمن الرحيم	Alhamd lilah rabi alealamin alrahman alrahim
إياك نعبد وإياك نستعين	'Iiaak naebud wa'iiaak nastaein
صراط الذين أنعمت عليهم	Sirat aladhin 'aneamt ealayhim
غير المغضوب عليهم ولا الضالين	Ghayr almaghdub ealayhim wala aldaaliyn
تراهم ركعاً سجداً يبتغون فضلاً من الله ورضوانا	Tarahum rakean sujdan yabtaghun fadlan min allah waridwana
وعد الله الذين آمنوا وعملوا الصالحات	Waead allah aladhin amnuu waeamiluu alsaalihat
مغفرة واجراً عظيما	Maghfirat wajraan eaziman
قل اعوذ برب الفلق من شر ما خلق	Qul aeuidh birabi alfalaq min shari ma khalaq
صلى الله على نبينا محمد وعلى آله وصحبه أجمعين	Salaa allah ealaa nabiina muhamad waealaa alah wasahbih 'ajmaein
سبحان الله وبحمده سبحان الله العظيم	Subhan allah wabihamdih subhan allah aleazim

Dataset Description

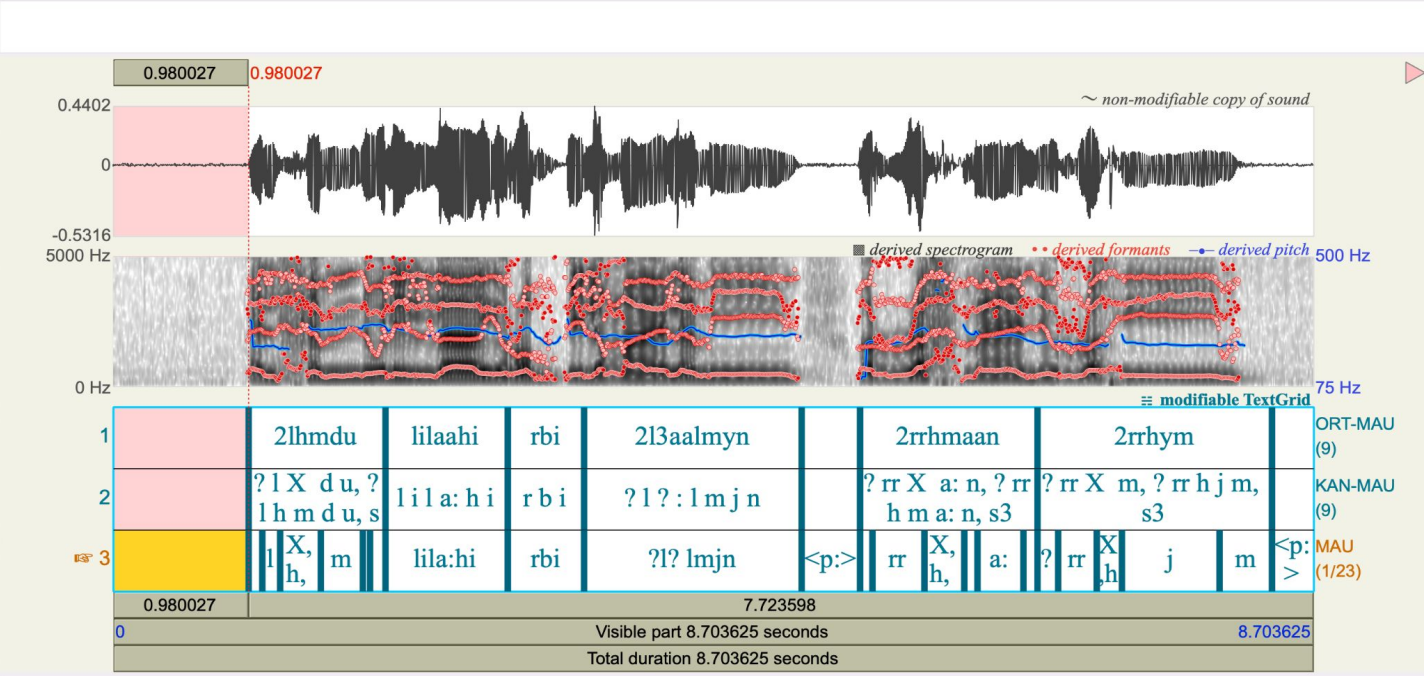
- Number of speakers

Native Arabic speakers	20 Male	20 Male
Non-native Arabic speakers	20 Female	20 Female
Total	40	40

- Non-native nationalities: Arabic language learners from central and west African countries.
- Number of sentences: Each speaker uttering 10 sentences 5 times.
- Number of phonemes: 26 phonemes.
- Number of audio files: 4k audio files.
- Average length of audio files: 6 seconds.

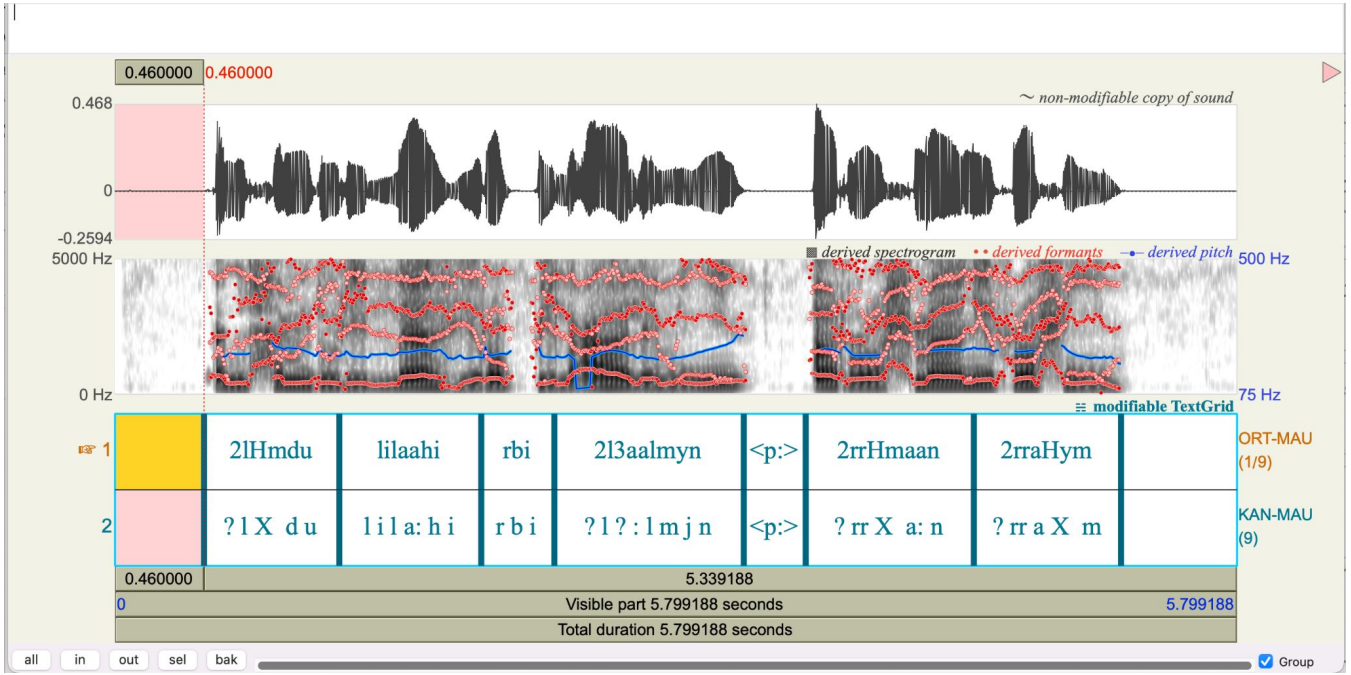
Annotation examples

- With pronunciation errors



Annotation examples

- Without errors



Training hyperparameter information

wav2vec2-xls-r-300m

- Num Epochs = 150
- Instantaneous batch size per device = 20
- Total train batch size (w. parallel, distributed & accumulation) = 80
- Adam Optimizer with default betas and epsilon
- Gradient Accumulation steps = 4
- Total optimization steps = 150
- Number of trainable parameters = 311255194
- Learning rate = 5e-4
- Lr_scheduler = “linear”
- Weight_decay = 0.001
- Warmup Steps = 250
- GPU Used = A100-SXM4-40GB

Performance Evaluation

- Final Train loss = 2.436
- Final Eval loss = 2.493
- Eval WER = 1.00
- Eval CER = 0.865
- Test WER = 1.00
- Test CER = 0.891

Predicted_string	Original_string
2iy 3yn	2iyaaka n3bdu wa 2iyaaka nst3yn

Challenges

1. Small data
2. Different phonetic transcription (ATR)
    a. Changing the phonetic transcription
3. Consider speech rate and intonation

Instructions

- Please edit directly on this google slide deck. During the presentation, you will use a provided laptop for the presentation.

- The final presentation should consist of 3 min presentation + 1-2min QA from judges. Please stick to the time as we will stop presentations that exceed 5 min.

- In your presentation please consider the following:

- Goal of the project and what social or economic impact could it create
- What it makes interesting and/or innovative ?
- Challenges you have overcome
- What have you learned from it ?
- What makes the project special or gives your proposal an edge over similar solutions in the market ?

TIPS and guidelines

- Please do not copy the contents from other materials (if it is very difficult to redraw, it is acceptable with the appropriate citation information).
- It depends on the audience, but it is a good idea to spend some time clearly presenting the introduction/motivation/problem setups
- Use a simple picture to emphasize your method/concept
- Long sentences in slides are not a good idea
- If you are showing numbers, please extract important numbers or highlight important numbers
- Add a take-home message in your final part

ASR problem

- There is no single ASR problem
  - Microphone: Close-mic, throat-mic, microphone array, audio-visual
  - Sources: band-limited, background noise, reverberation
  - Speaker: speaker dependent, speaker independent
  - Language: open/closed vocabulary, vocabulary size, read/spontaneous speech
  - Output: Transcription, speaker id, keywords