Speech-to-Speech Translation for Arabic Dialects

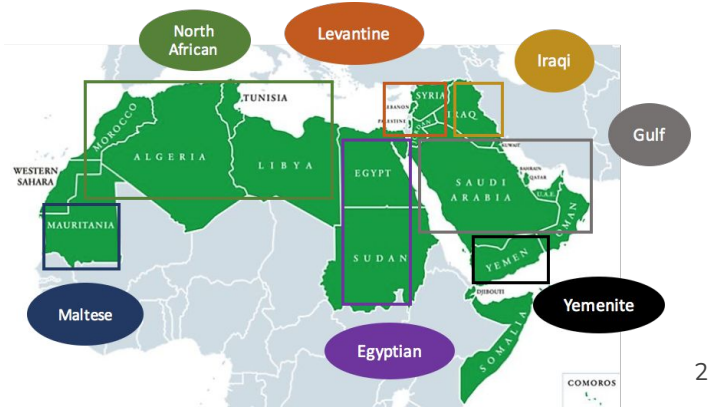
Hawau Toyin - MBZUAI, UAE
Nizar Youssef - American School, Qatar
Fatemah Husain - Kuwait University, Kuwait

Mentor: Harshita Diddee

SLT 2022

Motivation

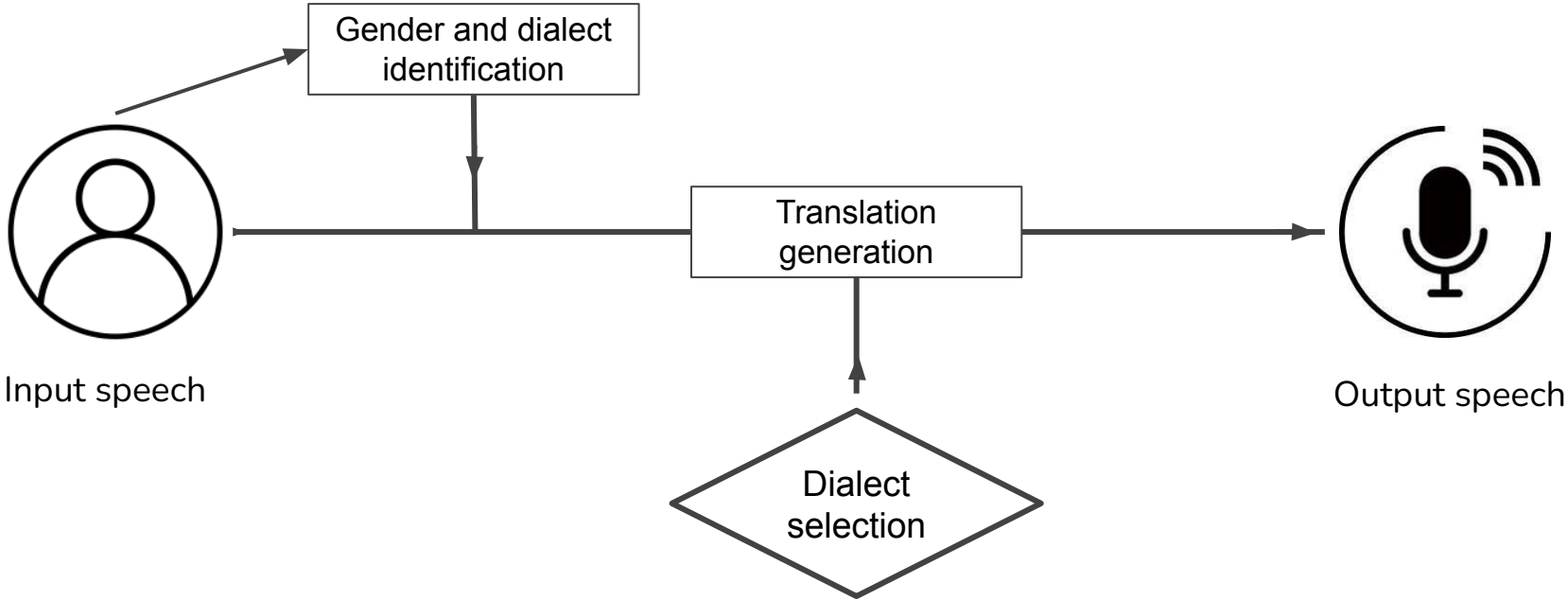
- Arabic has been characterized by **diglossia** since the 7th century [1].
- The opposition is between the Modern Standard Arabic (MSA), used and learned in formal settings and **diverse dialects that serve as Arabic mother tongues**.
- In any of the GCC countries, **more than 50% of the workforce consist of non-native arabic speakers** migrant workers, even if they learn Arabic, conversing with native arabic speakers will still be challenging [2].
- There are **more than 7 Arabic dialects with large variations** [3], thus, developing a system that can work with all dialects in addition to MSA and English is very important even for Arabic native speakers.



Objectives

- To convert speech from English to MSA or dialectal Arabic, and dialectal Arabic translation.
- Improve Automatic Speech Recognition (ASR) and speech Machine Translation (MT) systems for Arabic dialects, while incorporating the speaker’s gender.

Proposed System



Novelty

- There is no published work or tool on speech MT of Arabic Dialects.
- It can create a foundation for research and projects focused on integrating gender to Arabic speech generation and translation.

Challenges

- No publicly available dataset for parallel Arabic dialects and English.
- Limited Arabic speech dialectal processing tools.
- Gender variations on Arabic text add complexity.
- Inconsistency among Arabic dialects.
- Limited time frame to work (**~2 days**)

Egyptian: Otta – أطة

Levantine: Bisse – بسة

Gulf: Qatwa – قطوة

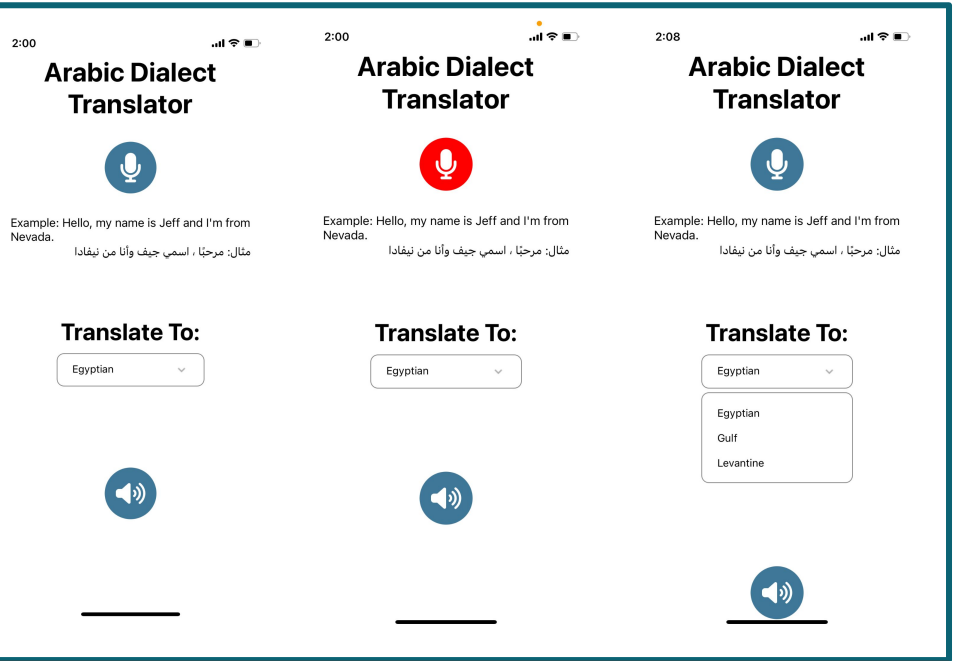
Moroccan: Qetta- قطة

Iraqi: Bazzuna – بزونة

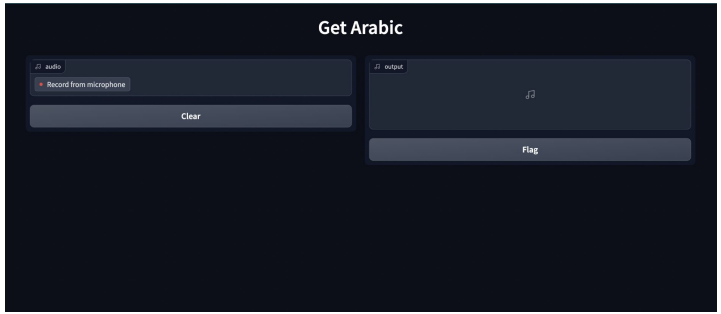
Yamani: Sanoura/ demah

سنورة / دمة –

Implementation - Current Progress



Mobile-based Interface



Web-based Application for Eng to MSA

Lessons Learned

- Limited speech resources in Arabic language creates a large gap in the available technological solution, which impacts researchers in the field and also members of its community.

Future Work

- Improve current implementation to include all proposed modules
- Create parallel Arabic dialectal, MSA and English Dataset

Q&A

Hawau Toyin - Mohamed Bin Zayed University of AI, UAE
hawau.toyin@mbzuai.ac.ae
Nizar Youssef - American School, Qatar
23nyoussef@asd.edu.qa
Fatemah Husain - Kuwait University, Kuwait
f.husain@ku.edu.kw

References

[1] Al-Busaidi, F. Y. (2015). Arabic in foreign language programmes: difficulties and Challenges. Journal of Educational and Psychological Studies-Sultan Qaboos University (Pages 701-717), 9(4).

[2] Kamrava, M.; Babar, Z. (2012). "Situating Labor Migration in the Persian Gulf". In Kamrava, M.; Babar, Z. (eds.). Migrant Labor in the Persian Gulf. New York: Columbia University Press., p. 8.

[3] Habash, N. Y. (2010). Introduction to Arabic natural language processing. Synthesis lectures on human language technologies, 3(1), 1-187.