A Romanization System and WebMAUS Aligner for Arabic Varieties

Jalal Al-Tamimi[*], Florian Schiel[**],
Ghada Khattab[†], Navdeep Sokhey[‡‡], Djegdjiga Amazouz[‡],
Abdulrahman Dallak[††], Hajar Moussa[⋆]

[*]**Université Paris Cité, CNRS, Laboratoire de linguistique formelle (LLF), F-75013 Paris France**
[**]Bavarian Archive for Speech Signals, Schellingstr. 3 80799 München Germany
[†]Newcastle University, Newcastle upon Tyne, NE1 7RU UK
[‡‡]Virginia Polytechnic Institute and State University, 220 Stanger Street, 331 Major Williams Hall (0225),
[‡‡]Laboratoire de Phonétique et Phonologie (LPP)-CNRS, Université Sorbonne Nouvelle, 75005, Paris France
[††]Newcastle University, Newcastle upon Tyne, NE1 7RU UK and Jazan University, Jazan, 45142, Saudi Arabia,
[⋆]King Abdul-Aziz University, Jeddah, 21589, Saudi Arabia.

Aims

- Arabic variety-independent romanization system
- Arabic variety-independent WebMAUS service
- ASR ⇒ Text-to-Phoneme alignment of pre-transcribed audio
- Free Phone Recognition of untranscribed audio to phones
- `https://clarin.phonetik.uni-muenchen.de/BASWebServices/`

Romanization System

- Arabic script for Standard Arabic transparent; less transparent for dialectal forms
- No common romanized system for Standard or dialectal Arabic (e.g., Buckwalter, Arabizi, etc.)
- Specific dictionary to deal with variants ⇒ Complex to design due to reductions, variation;
- Unavailable “phonemes” in Standard form, e.g., /zˤ/ for ”ظ“ and/or Variants in pronunciation, e.g., /x vs χ/ for ”خ“ or /ɣ vs ʁ/ for ”غ“
- Phonetically-based orthographic transcription of spoken speech; **1 sound = 1 symbol**; ASCII characters
- **Al-Tamimi Romanization system; ATR** ⇒ Short vs long consonants and vowels represented with separate symbols; 98 symbols
- /x/ = “x”; /χ/ = “X”; variants of /q/ ⇒ /ʔ/ = “2”, /q/ = “q”, /g/ = “g”; /sˤsˤ/ = ”SS”; /ɪ/ = “i”; /eː/ = “ee”, etc.

WebMAUS services

- 22 speech and language processing tools; 42 languages and language varieties ⇒ 1243 trained acoustic models, using HTK
- General or Pipeline framework ⇒ several individual services, e.g., ASR, G2P, Forced-alignment, etc...
- Arabic WebMAUS service ⇒ released 2021-12-17; version 2 on 2022-04-15
- Training set ⇒ 6610 recordings, 5 dialects; 94 speakers; total duration of 16h10min and 509804 labelled phone segments.
- Acoustic Model ⇒ posterior probability for a phone class given a segment of speech
- Total 105 models ⇒ 69 available phonemes; 16 cloned from other languages; 13 geminates from singletons; 7 noises, pauses, etc
- ATR ⇒ X-SAMPA (G2P) ⇒ Forced Alignment; WebMINNI (Free Phone Recognition)

Examples - Jeddah Arabic - Alignment (Fig. 1 & 3); WebMINNI (Fig. 2 & 3)

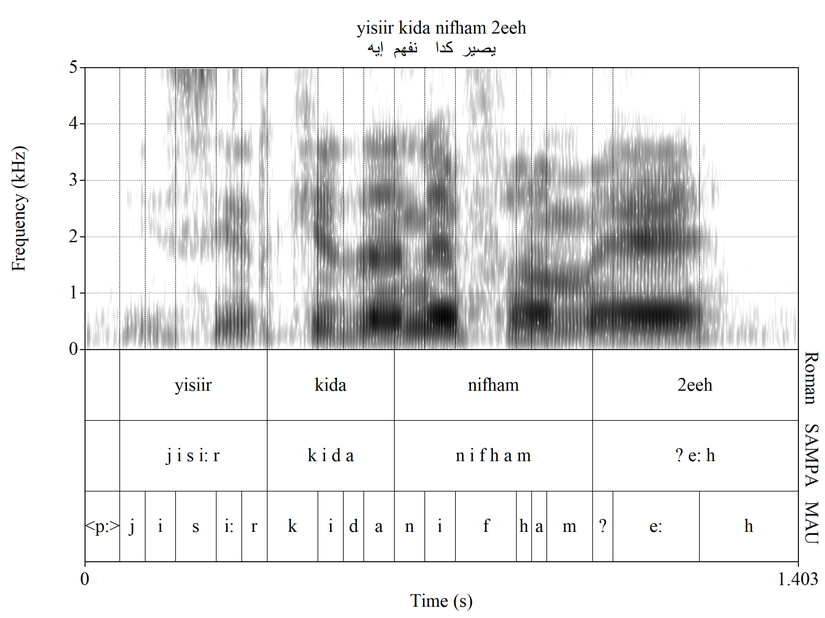

Figure 1. Alignment of “How do we understand this”

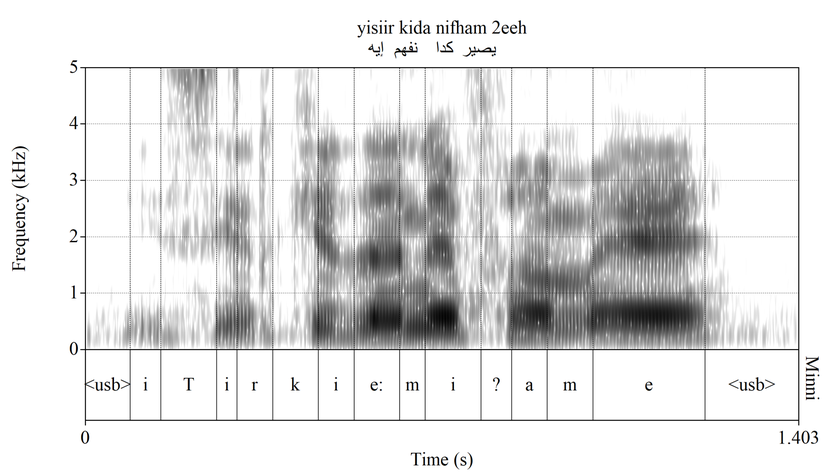

Figure 2. WebMINNI output of file “How do we understand this”

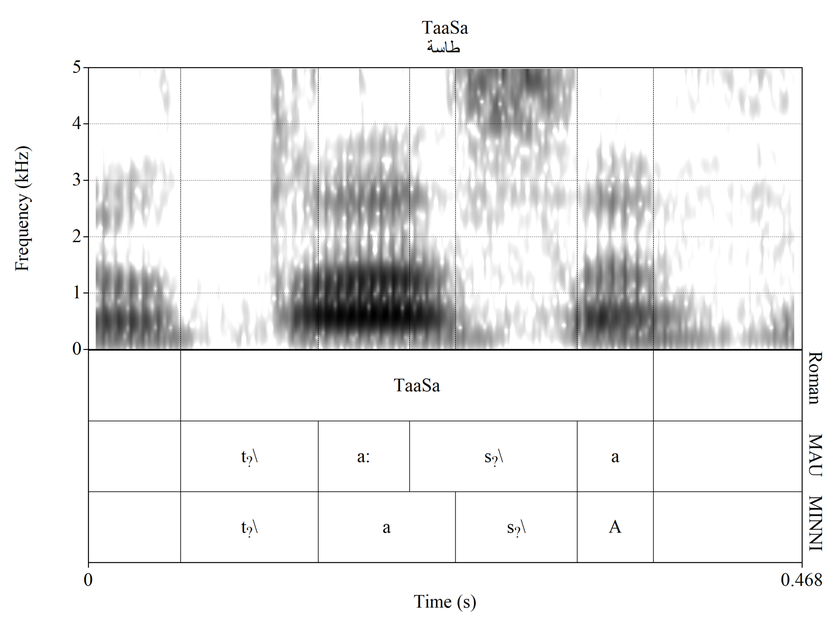

Figure 3. Alignment of word “cup” (tier 2); WebMINNI of same word (tier 3)

Conclusion and next steps

Variable success rates; “gutturals” and geminates well-identified; future: development of region-specific WebMAUS services. Facilitates transcription and alignment of data; support large-scale research on Arabic dialects