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.
[*]jalal.al-tamimi@u-paris.fr, [**]schiel@bas.uni-muenchen.de,
[†]ghada.khattab@newcastle.ac.uk, [‡‡]navsokhey@vt.edu, [‡]djegdjiga.amazouz@sorbonne-nouvelle.fr,
[††]A.Dallak2@newcastle.ac.uk/[††]adallak@jazanu.edu.sa, [⋆]hmoussa@kau.edu.sa

Abstract

This paper presents the results of an ongoing collaboration to develop an Arabic variety-independent romanization system that aims to homogenize and simplify the romanization of the Arabic script, and introduces an Arabic variety-independent WebMAUS service offering a free to use forced-alignment service fully integrated within the WebMAUS services. We present the rationale for developing such a system, highlighting the need for a detailed romanization system with graphemes corresponding to the phonemic short and long vowels/consonants in Arabic varieties. We describe how the acoustic model was created, followed by several hands-on recipes for applying the forced alignment webservice either online or programatically. Finally, we discuss some of the issues we faced during the development of the system.

Keywords: Arabic, romanization, dialect-independent, WebMAUS, aligner

1. Introduction and Motivation

The WebMAUS[1] is a suite of webservices that is free for academic users that – at the time of writing – comprises 22 speech and language processing tools (Kisler et al., 2017). Since its introduction in 2013, roughly 17 million media files from users from all over the world have been processed by this service suite. The core component around which most of the services are grouped is the well known WebMAUS speech segmentation and labelling engine first introduced in 1999 (Schiel, 1999). A powerful pipeline framework allows the concatenation of several individual services; for instance, automatic phonetic and syllabic segmentation and labelling of a speech recording is first performed using Automatic Speech Recognition (ASR) followed by a Text-to-phoneme translation, the WebMAUS engine and a Syllabification service in one processing call. Another point that distinguishes the WebMAUS services is the fact that users can choose between a sophisticated web interface or a programmatic access to each service (and the pipeline) using REST based synchronous service calls[2].

Currently, the WebMAUS services process 42 languages and language varieties, and utilize 1243 trained acoustic phone models based on HTK (Young et al., 2015). Up to now, Arabic varieties have not been part of the supported languages. The reasons are two-fold: firstly, there is a general lack of sufficient phonetically annotated speech material for each of the Arabic varieties, and secondly, Arabic orthography has a very complex relationship with the phonology of the individual Arabic variety. While diacritised Arabic orthography is transparent, using orthography to represent individual Arabic varieties (rather than Standard Arabic) is a relatively recent practice that is much more complex and that can be a lot less transparent. Nevertheless, there is quite a demand to provide Arabic speech processing services, as can be seen for instance in Google's Cloud Speech-to-text system that alone provides individual ASR models for no less than 15 Arabic varieties[3], but with an output that is heavily dependent on the Classical Arabic orthography.

This paper presents the results of an ongoing col-

---

[1]https://hdl.handle.net/11858/00-1779-0000-0028-421B-4, also known as "BAS WebServices"

[2]See https://clarin.phonetik.uni-muenchen.de/BASWebServices/services/ help for details

[3]United Arab Emirates, Bahrain, Algeria, Egypt, Israel, Iraq, Jordan, Kuwait, Lebanon, Morocco, Oman, State of Palestine, Qatar, Saudi Arabia, Tunisia.

laboration between researchers from various institutions led by the first author (currently with the Laboratoire de Linguistique Formelle at the Université Paris Cité & CNRS) in collaboration with various researchers (co-authors) and researchers from the Bavarian Archive for Speech Signals (BAS) at the University of Munich, Germany, led by the second author. This collaboration allowed the emergence of a first **Arabic variety-independent WebMAUS service**. The fundamental conceptual idea of this endeavour was that we did not start from a Classical Arabic[4] script as processing input (which would necessarily lead to the aforementioned problems of orthography-phonology relationships) and the reliance on a lexical-dependent language dictionary that spans all Arabic varieties. Rather we developed an **Arabic variety-independent romanization system** that allows the easy orthographic transcription of nearly all Arabic varieties; we name this phonemically-based romanization system after its creator **Al-Tamimi Romanization** (ATR, section 2). In this paper we describe the chosen "phoneme" system that aims to cover the phonemes of nearly all Arabic varieties, and the data collection and training of the corresponding acoustic WebMAUS models (section 3).

The resulting Arabic WebMAUS services were released to the public on 2021-12-17 (version 1), since updated to include additional resources (version 2, on 2022-04-15). In section 4 we present easy to follow step-by-step instructions as well as links to suitable example recordings which allow the reader to test the new system.

2. Al-Tamimi Romanization of Arabic Orthography

2.1. Rationale for a new romanization system

It is well known that Arabic script provides transcriptions of only consonants and long vowels; vowelisations (or diacritisation, i.e., adding small vowel diacritics on top/bottom) of the short vowels is generally optional because it can be predictable based on the utterance meaning. Unfortunately, this can create a barrier for any forced-alignment system to work properly, unless a specific dictionary is available that allows for a specific lexical item to be associated with a specific phonemic transcription that is then used by the forced-

---

[4]Arabic is a diglossic language whereby, until recently, Classical and Standard varieties have been used in writing while dialectal varieties have mainly been used in spoken discourse. Technological advances and the rise of social media have, however, led to a significant increase in using both Arabic orthography and multiple Romanization systems to represent dialects in writing

alignment system to match the Acoustic Model (AM) to a phoneme. When looking at Arabic varieties, this adds another barrier due to the fact that nearly all Arabic varieties are spoken-only varieties with no specific written script available (except in some regions). Indeed, there is no standardized writing system across spoken varieties of Arabic. This variation of spoken Arabic has increased with the popularization of online writing and texting. While some shared spelling conventions exist across these platforms (in some speech communities more so than others), there may be multiple spelling variations accepted for a particular word item. This is unlike spelling conventions in Modern Standard Arabic, which are highly standardized. And even with this, it is clear that even Standard Arabic lacks a common and a standardized Romanization system, when one looks at the number of systems available online.

Finally, when looking at current systems, including Google's Cloud Speech-to-text system, it only provides Arabic script output for Arabic varieties and does not include vowelisations. Indeed, many researchers trying to use large-scale datasets face massive issues in dealing with the scarcity of open-source force-alignment systems for Arabic that are able to provide phonetic and phonological level transcription and segmentation of speech outputs. Such a restriction prohibits the use and expansion of research on Arabic.

2.2. Details of ATR

Our aim was then to develop a specific transliteration system that provided a transparent and direct match between sounds and orthography with a 1-to-1 match between a produced sound and a symbol to transcribe it, using ASCII characters. We decided to adopt such a system as it is transparent and allows researchers to provide a phonetically-based orthographic transcription of spoken speech. Indeed, other systems exist, e.g., Arabizi (or the system used for sms and chats), but also other transliteration systems, e.g., Buckwalter Arabic translator of Arabic script to romanized symbols. The issues with the first is its generic nature in sometimes combining multiple sounds produced to the same symbol, e.g., the symbols '2' and 'a' for the letter hamza (for IPA **ʔ**; X-SAMPA '?'), and specifically, which target sound the transliteration is trying to represent. For the latter, and although it can allow for vowelisation of short vowels, these are unfortunately rarely transcribed in the orthographic transcriptions. If we look at the Arabic datasets within the LDC database, we observe that the majority of transcribed datasets are with an Arabic script and without vowelisation. This adds a major barrier to any large-scale phonetic, phonological and morphological analysis of the Arabic

sound system from standard and/or dialectal spoken corpora.

2.3. Types of phonemes used

ATR contains 98 phonemes that cover all possible sounds present in the various varieties investigated in the teams' research, from the phonetic-phonological analysis of gemination (doubling of consonants) in Lebanese Arabic (Al-Tamimi and Khattab, 2011; Al-Tamimi and Khattab, 2015; Al-Tamimi and Khattab, 2018; Khattab and Al-Tamimi, 2014), to the phonetic-phonological analysis of pharyngealized and guttural consonants in Levantine Arabic (Al-Tamimi, 2017; Al-Tamimi, 2021), automatic speech processing of code-switching in Algerian Arabic and French (Amazouz, 2019; Amazouz et al., 2018), the linguistic accommodation in Bahraini Arabic (Sokhey, 2019; Sokhey, Forthcoming), feature specification of the voicing contrast in Jizani Arabic (Dallak, forthcoming), intonational studies of Arabic dialects, e.g., Saudi Arabic (Moussa, 2020; Moussa and Al-Tamimi, 2018), and finally, child-directed speech in Lebanese Arabic (Khattab and Al-Tamimi, 2013; Khattab and Al-Tamimi, 2015). The full list of transliterated symbols is available in Tables 1 and 2, for long and short phonemes, respectively (see column ATR System). The motivation of having such variety of sounds was to allow for any type of production to be transcribed and documented. As can be seen, some sounds used are dialect-specific, e.g., /zˤ/ or /zˤzˤ/ that are available in some Levantine, Jeddah, Mekkah and Medina Arabic and in many Western North African dialects; or the sound /ɡ/ as a reflex of the classical Arabic sound /q/ found in many dialects or as a dialectal reflex of /ʒ/ seen in Egyptian Arabic.

Looking specifically at this system, we present two sets of sounds: the "long" (see Table 1) and the "short" sounds (see Table 2). By "long" and "short" we mean, geminate consonants and long vowels for the long phonemes set and singleton consonants and short vowels for the short phonemes set. It was important to create different sets for each of the "long" and "short" phonemes to allow for the AM to be able to deal with the systematic acoustic differences observed across the types of phonemes. Indeed, previous research showed that short and long vowels have acoustic differences on both temporal and non-temporal domains, with shorter durations and more reduced vowel spaces in the "short" set, in comparison with the "long" set that is more peripheral (Al-Tamimi, 2007a; Al-Tamimi, 2007b; Al-Tamimi and Ferragne, 2005). The same can be said for the differences between singleton and geminate consonants, as previous research showed clear temporal and non-temporal differences between the two sets, with geminates

being associated with longer consonantal durations and non-temporal differences associated with a stronger/tense production, and with specific variations in the surrounding vowels (Al-Tamimi and Khattab, 2011; Al-Tamimi and Khattab, 2015; Al-Tamimi and Khattab, 2018; Khattab and Al-Tamimi, 2014). Hence one can see Table 1 as a mirror of Table 2, but with different grapheme forms and with different acoustic models (see below).

IPA	ATR System	X-SAMPA
ʔʔ	22	??
bb	bb	bb
tt	tt	tt
θθ	t\t\	TT
ʒʒ	jj	ZZ
ħħ	HH	X\
xx	xx	xx
χχ	XX	XX
dd	dd	dd
ðð	d\d\	DD
rr	rr	rr
zz	zz	zz
ss	ss	ss
ʃʃ	s\s\	SS
sˤsˤ	SS	s_?\s_?\
dˤdˤ	DD	d_?\d_?\
tˤtˤ	TT	t_?\t_?\
ðˤðˤ	D\D\	D_?\D_?\
zˤzˤ	ZZ	z_?\z_?\
lˤlˤ	LL	l_?\l_?\
ʕʕ	33	?\
ɣɣ	GG	GG
ʁʁ	G\G\	G\G\
ff	ff	ff
qq	qq	qq
ɡɡ	gg	gg
kk	kk	kk
ll	ll	ll
mm	mm	mm
nn	nn	nn
hh	hh	hh
ww	ww	ww
jj	yy	jj
t͡ʃt͡ʃ	chch	tStS
d͡ʒd͡ʒ	djdj	dZdZ
vv	vv	vv
pp	pp	pp
iː	ii	i:
ɪː	II	I:
eː	ee	e:
ɛː	EE	E:
æː	aeae	{:
aː	aa	a:
ɑː	AA	A:
ɔː	OO	O:
oː	oo	o:
uː	uu	u:
ʊː	UU	U:
əː	@@	@:

Table 1: List of long phonemes used in the ATR system, with IPA in Unicode, romanization system and IPA in X-SAMPA

IPA	ATR System	X-SAMPA
ʔ	2	?
b	b	b
t	t	t
θ	t\	T
ʒ	j	Z
ħ	H	X\
x	x	x
χ	X	X
d	d	d
ð	d\	D
r	r	r
z	z	z
s	s	s
ʃ	s\	S
sˤ	S	s_?\
dˤ	D	d_?\
tˤ	T	t_?\
ðˤ	D\	D_?\
zˤ	z\	z_?\
lˤ	L	l_?\
ʕ	3	?\
ɣ	G	G
ʁ	G\	G\
f	f	f
q	q	q
g	g	g
k	k	k
l	l	l
m	m	m
n	n	n
h	h	h
w	w	w
j	y	j
tʃ	ch	tS
dʒ	dj	dZ
v	v	v
p	p	p
i	i	i
ɪ	I	I
e	e	e
ɛ	E	E
æ	ae	{
a	a	a
ɑ	A	A
ɔ	O	O
o	o	o
u	u	u
ʊ	U	U
ə	@	@

Table 2: List of short phonemes used in the ATR system, with IPA in Unicode, romanization system and IPA in X-SAMPA

3. Arabic WebMAUS

The WebMAUS technique (Schiel, 1999) uses two machine-learned models: the AM that estimates the posterior probability for a phone class given a segment of speech, and a pronunciation model (PM) that estimates the probability of a sequence of spoken phones. In this paper we concentrate on the creation of the AM; the PM will be subject to future work of our on-going collaboration.

The WebMAUS AM needed to cover all 98 phoneme classes described in the previous section to enable the system to represent nearly all Arabic varieties. Ideally, this would have required a verified segmented and labelled training set of speech recordings comprising enough samples of each phoneme class from every Arabic variety in which the phoneme is found and spoken by at least 50 native speakers of both sexes. Since there are no publicly available resources to fulfil these requirements, we collected speech recordings from a variety of ongoing scientific projects of various Arabic varieties (see above). These varieties are: Bahraini, Saudi Arabian, Lebanese, Levantine (comprised of Lebanese, Syrian, Palestinian Arabic). Within our collaboration, the recordings and their accompanying transcriptions were collected, unified and merged into a common annotation format, and first automatically segmented using the language-independent system of WebMAUS then manually verified to obtain a speech signal with time-aligned orthographic/phonetic transliteration and segmentation. All transcribers who worked on the corpora were trained phoneticians who were also L1 speakers of the dialects in question. The transcription convention that was followed was informed by a broad phonetic transcription of the incoming signal which could be accommodated within the ATR system. This was deemed optimal as it allowed us to adopt a bottom-up approach to the transcription which relied on what speakers said and how they said it; this was better for training the AM and was not hindered by the lack of published comprehensive sources on the phonetic and phonological patterns of the varieties in question. Finally, these were used as input to train the AM. The current training set in version 2 comprises 6610 recordings, from 94 speakers, with a total duration of 16h10min and 509804 labelled phone segments.

Problems arose for 29 phoneme classes (see Table 3) that were either not present in the merged corpus or under-represented (sample size less than 75 samples). To solve this problem we cloned 16 corresponding acoustic models from other languages and – as a last resort – cloned 13 existing non-geminate models as geminates.[5] Finally, we added 7 silence and noise models from other languages to the set to enable the modelling of non-speech, speech noise (such as coughs, breathing, laughter, etc.) and non-speech background noise.

The resulting 105 models were merged into a language MAUS parameter set[6] labelled with

---

[5] The cloning of these geminates from simplex phone models appears to be a rather drastic measure, since now the system is unable to acoustically distinguish between them. But in practical terms concurring acoustic geminate and non-geminate models never occur in MAUS because they are not allophonic variants of each other.

[6] MAUS version 5.105 and higher.

@: – ə	{ – æ	{: – æː
D_?\D_?\– ðˤðˤ	DD – ðð	dZdZ – ʤʤ
E – ɛ	E: – ɛː	G\– ʁ
gg – gg	G\G\– ʁʁ	I – ɪ
I: – ɪː	l_?\l_?\– lˤlˤ	O – ɔ
O: – ɔː	p – p	pp – pp
qq – qq	tStS / ʧʧ	TT – θθ
U – ʊ	U: – ʊː	v – v
vv – vv	X – χ	XX – χχ
z_?\z_?\– zˤzˤ	d_?\d_?\– dˤdˤ	

Table 3: Un-trainable acoustic models in both X-SAMPA and IPA Unicode

RFC5646 language code "arb", i.e. just the iso639 code without any following region code(s), to indicate that this is not a specific Arabic variety but rather a macro language model suitable for nearly all Arabic varieties.

Additionally, we calculated a simple bi-gram model based on the phone label sequences found in the training set and set up a free phone recognition service WebMINNI for Arabic varieties.[7]

4. How to Use the Service

There are several ways to apply the new Arabic WebMAUS services to recordings of Arabic varieties. In the following we will give easy to follow step-by-step instructions for usage through the web interface as well as one example for a programmatic REST call.

In all examples we will use an ATR transcript file and the associated recording of a speaker who speaks the Urban Hijazi variety as spoken in Jeddah that was not part of the training data. The examples can be downloaded from here:

`https://www.bas.uni-muenchen.de/Bas/BASWebServices/DOCS/Arabic/example.txt`

`https://www.bas.uni-muenchen.de/Bas/BASWebServices/DOCS/Arabic/example.wav`

Before getting started with the exercises below please download these `example` files to your local computer.

4.1. Grapheme-to-Phoneme Conversion

The service "G2P" reads an ATR transcript stored in a UTF-8 plain text file, tokenizes the text into words, removes punctuation and converts the romanization into X-SAMPA. This can be a first step to obtain phonetic transcriptions of a transcript that uses our ATR system. The service offers a number of output formats but for this demonstration we will use the BAS Partitur Format (BPF)[8] that all WebMAUS services can read.

- go to: `https://hdl.handle.net/11858/00-1779-0000-0028-421B-4`
- in the left-hand pull down menu "Show service side bar" select service "G2P"
- drag&drop the file `example.txt` in the file drop area
- click the "Upload" button; the log window below should turn green
- select "Arabic (macro)" from the "Language" pull down menu
- select "true" from the "Keep annotation markers" pull down menu. This causes the webservice to 'protect' the silence marker `<p:>` against translation into X-SAMPA.
- select "maus" from the "Tool embedding" pull down menu
- confirm the "Terms of Usage" and click on "Run Web Service"; a link to a result file `example.par` should appear, and the log window below should remain green
- click on the BPF result file to inspect it; it should contain two tier sections, 'ORT' and 'KAN' listing the word tokens from the input text file (ORT) and the canonical pronunciation encoded in X-SAMPA (KAN).
- left-click on the result file link to download it to your local computer

4.2. MAUS Alignment with BPF input

The service "WebMAUS General" reads a word-tokenized canonical transcript in either X-SAMPA or IPA and performs a MAUS segmentation. This transcript can be based on step 1 above, i.e., an automated Grapheme-to-Phoneme conversion or a preexisting IPA/X-SAMPA transcript (see below). This is the first step towards obtaining a forced-alignment of a pretranscribed transcript. The service offers a number of output formats but for this demonstration we will use praat TextGrid output format because this can be displayed within the webbrowser using the "Emu webapp" viewer (Winkelmann et al., 2017).

- if not already there, go to: `https://hdl.handle.net/11858/00-1779-0000-0028-421B-4`
- in the left-hand pull down menu "Show service side bar" select service "WebMAUS General"
- drag&drop the file `example.wav` and the result file `example.par` from the previous exercise in the file drop area

---

[7]`https://clarin.phonetik.uni-muenchen.de/BASWebServices/interface/WebMINNI`

[8]`https://www.bas.uni-muenchen.de/forschung/Bas/BasFormatseng.html`

- click the "Upload" button; the log window below should turn green
- select "Arabic (macro)" from the "Language" pull down menu
- confirm the "Terms of Usage" and click on "Run Web Service"; a link to a result file `example.TextGrid` should appear, and the log window below should remain green
- click on the TextGrid result file to inspect it; click on the blue "Segmentation Preview" logo in the popup window; a new tab should open in your browser displaying the MAUS segmentation in words, canonical word pronunciation and phones

Instead of using the G2P result as input to this service, you might create your own phonetic input transcript in the "KAN" section of the input BPF file[9]; this can be handy if you have no ATR transcript input but rather a phonetic transcript in IPA or X-SAMPA. By clicking on the "Show inventory" button right of the "Language" option you can see a table with all supported phonetic symbols for Arabic varieties.

The next exercise shows how to perform both a "G2P" and a "WebMAUS General" (and possibly other services) in one step.

4.3. Pipeline with ATR input

This exercise demonstrates how to obtain a forced-alignment output in a Praat's native TextGrid format from an ATR transcription. The service "Pipeline without ASR" reads an ATR transcript stored in a UTF-8 plain text file, tokenizes the text into words, removes punctuation, translates the romanization into X-SAMPA, and then performs a MAUS segmentation. As you will note this service allows to perform several BAS WebServices after another (in a pipeline); by selecting a different pipeline name you will get different results accordingly.

- if not already there, go to: `https://hdl.handle.net/11858/00-1779-0000-0028-421B-4`
- in the left-hand pull down menu "Show service side bar" select service "Pipeline without ASR"
- drag&drop the file pair `example.txt` and `example.wav` in the file drop area
- click the "Upload" button; the log window below should turn green
- select "G2P_MAUS" from the "Pipeline name" pull down menu[10]

---

[9]Or you can specify the input X-SAMPA transcript in a CSV table

[10]Exactly the same pipeline is performed in the "WebMAUS Basic" service but without any expert options.

- select "Arabic (macro)" from the "Language" pull down menu
- confirm the "Terms of Usage" and click on "Run Web Service"; a link to a result file `example.TextGrid` should appear, and the log window below should remain green
- click on the TextGrid result file to inspect it; click on the blue "Segmentation Preview" logo in the popup window; a new tab should open in your browser displaying the MAUS segmentation in words, canonical word pronunciation, and phones

Note that the "Pipeline without ASR" service offers more than the individual services in the pipeline; for instance you can use any text format as input, not just UTF-8 ASCII, or you may apply a number of signal processing steps before feeding the recording into the pipeline (see "TEXTENHANCE" and "AUDIOENHANCE" expert options).

4.4. Free Phone Recognition

The service "WebMINNI" segments and labels the Arabic input speech signal into phones without any input transcript. The algorithm is based on an HTK recogniser using a phone bigram as language model. Note that this recogniser does not distinguish between different Arabic varieties, and thus applies all 98 possible Arabic phoneme classes. This leads to slightly worse results than in other langues with much fewer phone classes (e.g. English with 34 classes).

- if not already there, go to: `https://hdl.handle.net/11858/00-1779-0000-0028-421B-4`
- in the left-hand pull down menu "Show service side bar" select service "WebMINNI"
- drag&drop the file `example.wav` in the file drop area
- click the "Upload" button; the log window below should turn green
- select "Arabic (macro)" from the "Language" pull down menu
- confirm the "Terms of Usage" and click on "Run Web Service"; a link to a result file `example.TextGrid` should appear, and the log window below should turn yellow because of a conversion warning that you can savely ignore

- click on the TextGrid result file to inspect it; click on the blue "Segmentation Preview" logo in the popup window; a new tab should open in your browser displaying the MAUS segmentation in phones

4.5. Programmatic webservice call

As mentioned above the BAS WebServices also offer a programmatic access to most services[11]. For example, the ELAN video labelling tool[12] uses this interface to the service "WebMAUS Basic" for the embedded automatic segmentation of transcribed speaker turns. To demonstrate this RESTful interface we will replicate the exercise 4.3 applying the "Pipeline without ASR" service.

The following example call can be issued from any Shell console assuming that the current directory contains the two files `example.wav` and `example.txt`[13]:

- start a UNIX command line tool such as "terminal/console"

- change the current directory to the directory where you have downloaded the `example` files:
  `cd <dir_with_examples>`

- for upload and service call copy & paste the following command[14]:
  ```
  curl -X POST
  -H "content-type:multipart/form-data"
  -F PIPE=G2P_MAUS
  -F SIGNAL=@example.wav
  -F TEXT=@example.txt
  -F LANGUAGE=arb
  -F OUTFORMAT=TextGrid
  "http://clarin.phonetik.uni-
  muenchen.de/BASWebServices/
  services/runPipeline"
  ```

  The (XML) answer should be something like[15]:

  ```
  <WebServiceResponseLink>
    <success>"true"</success>
    <downloadLink>
      (URL to the result file
         example.TextGrid)
    </downloadLink>
    <output></output>
    <warnings></warnings>
  </WebServiceResponseLink>}
  ```

- download the result file `example.TextGrid`:
  `curl -O (URL to the result file example.TextGrid)`

The result file should be identical to the TextGrid result file you produced in exercise 4.3.

5. Conclusion and Future Work

This paper presented our first efforts in developing a freely available 1) Arabic variety-independent romanization system that aims to homogenize and simplify the romanization of the Arabic script with all possible variants, from short to long vowels and consonants and 2) Arabic variety-independent WebMAUS service that allows for either the forced-alignment of pre-transcribed Arabic script using (or not) the romanization system or a forced-alignment blind system (aka WebMINNI). The rates of success of the current system are variable, with some sound categories (e.g., geminates, gutturals, nasals) been automatically segmented with an almost perfect rate (close to 95%), when compared to a manually segmented speech, while in other cases, the rates are variable. Indeed, this difference is expected, however, this can be used as a first step in data segmentation followed by manual corrections, when needed. The WebMINNI's performance is variable and again, some instances are perfectly identified (e.g., gutturals, pharyngealized and geminates). The system provides promising results and additional data from a variety of Arabic dialects will increase accuracy of the model. This will allow for large-scale research on the Arabic sound system to emerge. Our next steps are to continue building and using additional corpora to enhance the acoustic model and build regional variety independent systems. These steps will strengthen the system further.

6. Acknowledgements

7. Bibliographical References

Al-Tamimi, J. and Ferragne, E. (2005). Does vowel space size depend on language vowel inventories? Evidence from two Arabic dialects and French. In *Interspeech*, pages 2465–2468.

Al-Tamimi, J. and Khattab, G. (2011). Multiple cues for the singleton-geminate contrast in Lebanese Arabic: Acoustic investigation of stops and fricatives. *Proceedings of the 17th ICPhS, Hong Kong, China.*

Al-Tamimi, J. and Khattab, G. (2015). Acoustic cue weighting in the singleton vs geminate contrast in Lebanese Arabic: The case of fricative consonants. *Journal of the Acoustical Society of America*, 138(1):344–360.

Al-Tamimi, J. and Khattab, G. (2018). Acoustic correlates of the voicing contrast in Lebanese Arabic singleton and geminate stops. *Journal of Phonetics, Invited manuscript for the special issue of Journal of Phonetics, "Marking 50 Years of Research on Voice Onset Time and the Voicing Contrast in the World's Languages" (eds., T. Cho, G. Docherty & D. Whalen)*, 71:306–325.

Al-Tamimi, J. (2007a). *Indices Dynamiques et Perception des Voyelles : Étude Translinguistique en Arabe Dialectal et en Français*. Ph.D. thesis, Université Lyon 2, Lyon, France.

Al-Tamimi, J. (2007b). Static and Dynamic cues in vowel production: A cross dialectal study in Jordanian and Moroccan Arabic. In *Proceedings of the 16th International Congress of Phonetic Sciences (ICPhS), Saarbrücken, Germany*, pages 541–544.

Al-Tamimi, J. (2017). Revisiting acoustic correlates of pharyngealization in Jordanian and Moroccan Arabic: Implications for formal representations. *Laboratory Phonology: Journal of the Association for Laboratory Phonology*, 8(1):1–40, November.

Al-Tamimi, J. (2021). Acoustic correlates of epilaryngeal constriction in Levantine Arabic guttural consonants. In *Proceedings of the 4th Phonetics and Phonology in Europe (PaPE), Barcelona, Spain, 21-23 June 2021*.

Amazouz, D., Adda-Decker, M., and Lamel, L. (2018). The French-Algerian Code-Switching Triggered audio corpus (FACST). page 8.

Amazouz, D. (2019). *Linguistic and Phonetic Investigations of French-Algerian Arabic Code-Switching:*. Ph.D. thesis.

Dallak, A. (forthcoming). *The Laryngeal Specification of Obstruents in Arabic.* phdthesis.

Khattab, G. and Al-Tamimi, J. (2013). Influence of geminate structure on early Arabic templatic patterns. In M Vihman et al., editors, *The Emergence of Phonology: Whole-word Approaches and Cross-linguistic Evidence*, pages 374–414. January.

Khattab, G. and Al-Tamimi, J. (2014). Geminate timing in Lebanese Arabic: The relationship between phonetic timing and phonological structure. *Journal of Laboratory Phonology*, 5(2):231–269, January.

Khattab, G. and Al-Tamimi, J. (2015). The acquisition of gemination in Lebanese Arabic Children (Article: 0870). In *The Scottish Consortium for ICPhS 2015 (Ed.), Proceedings of the 18th International Congress of Phonetic Sciences. Glasgow, UK: The University of Glasgow. ISBN 978-0-85261-941-4.*

Kisler, T., Reichel, U., and Schiel, F. (2017). Multilingual processing of speech via web services. *Computer Speech & Language*, 45:326–347, September.

Moussa, H. and Al-Tamimi, J. (2018). Focus and F0 patterns in Jeddah Arabic. In *Proceedings of the BAAP (British Association of Academic Phoneticians), Kent University, UK*.

Moussa, H. A. (2020). *Jeddah Arabic Intonation: An Autosegmental-metrical Approach*. Ph.D. thesis, Newcastle University.

Schiel, F. (1999). Automatic Phonetic Transcription of Non-Prompted Speech. In *Proc. of the ICPhS*, pages 607–610, San Francisco, August.

Sokhey, N. (2019). *Vowel-Rounding in Bahrain: An Interactional Study on Accommodation Based on Perceived Interlocutor Identity*. Ph.D. thesis.

Sokhey, N. (Forthcoming). Vowel-Rounding in Bahrain: An accommodation study based on perceived interlocutor identity. In James Dickins, editor, *Voices of Arabia: Studies on the Languages of the Peninsula*. Routledge.

Winkelmann, R., Harrington, J., and Jänsch, K. (2017). EMU-SDMS: Advanced speech database management and analysis in R. *Computer Speech & Language*, 45:392–410, September.

Young, S., Evermann, G., Gales, M., Hain, T., Kershaw, D., Liu, X., Moore, G., Odell, J., Ollason, D., Povey, D., Ragni, A., Valtchev, V., Woodland, P., and Zhang, C. (2015). *The HTK Book (version 3.5a)*. 12.