



Forced Alignment and Pronunciation Variation in Low-Resource Corpora

YANKI ÖZTÜRK & MATTHIAS DREWS

Project definition & tasks

- ★ Use forced alignment to generate phoneme or word-level alignments
- ★ Analyze pronunciation variation by speaker group or region
- ★ Evaluate alignment quality manually or via proxy metrics (e.g., duration distributions)

Project focus

- ★ Due to limited resources, we put the focus on one low-resource language
- ★ First language chosen: Haitian Creole
 - Discarded due to insufficient data
- ★ Final language: Tatar
 - ~ 5 million speakers, 4 million of those native L1
 - Spoken predominantly in Tatarstan, a republic of Russia in the Volga Region
 - Turkic language of the Kipchak sub-branch
 - Used to be written in Arabic script, now in (modified) Cyrillic

Montreal Forced Aligner

- ★ Forced alignment: a technique to take an *orthographic transcription* of an audio file and generate a *time-aligned version* using a pronunciation dictionary to look up phones for words
- ★ Four primary passes of training:
monophone → triphone → feature transformation → speaker differences
- ★ MFA: set of tools to make forced alignment easier - we did not train or own acoustic model or create a custom pronunciation dictionary from scratch

```
mfa model download acoustic tatar  
mfa model download dictionary tatar
```

Subset (124)

tt - 47.8k rows

Split (6)

train - 9.32k rows

Search this dataset

client_id	path	audio	sentence	up_votes	down_votes	age	gender	accent	locale	segment	variant
string - classes	string - lengths	audio - duration (s)	string - lengths	int64	int64	string - classes	string - classes	string - classes	string - classes	string - classes	string - classes
2 values	39	1.11 9.7	5 120	2 3	0 1	1 value	1 value	1 value	1 value	1 value	1 value
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343458.mp3	0:00 / 0:03	Баш естө, хан хәзрәтләре.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343449.mp3	0:00 / 0:08	Төшөндө биек койналар белән әйләндереп алынган бер бакчада би...	3	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343441.mp3	0:00 / 0:06	Балык элөккән көнне тамаклары тук, балык элөкмәгән көнне тамаклары ач...	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343443.mp3	0:00 / 0:02	Бу - минем фикер.	2	1	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343444.mp3	0:00 / 0:05	Ашап туйгач, савытларында ашап бетерә алмаган сөтләре кала.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343445.mp3	0:00 / 0:04	Татар түгел динә - башың ватернын!	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343446.mp3	0:00 / 0:04	Бер көн бара, ике көн бара.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343447.mp3	0:00 / 0:03	Барысы да тынып калды.	3	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343449.mp3	0:00 / 0:03	Теге аждаһалы елгага кайтып жүтө.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343450.mp3	0:00 / 0:03	Бер алаша ала да чыгып китө.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343451.mp3	0:00 / 0:02	Барысы да китөләр.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343452.mp3	0:00 / 0:04	Кызык сөзгә, яшь балалар, кышкы китчө!	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343455.mp3	0:00 / 0:06	Сафа Гәрәй хан хатынларым, балаларым, якынларым алып Болгар капкасынан...	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343556.mp3	0:00 / 0:04	Беренчедән, сөз миңа түгел, миң сөзгә рәхмәтләр.	3	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343557.mp3	0:00 / 0:03	Бүген ял көне.	2	0	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad ec05d6e6304d568ab7063daf915d3fccfb4dd44b81646bcd13...	tt_train_0/common_voice_tt_17343559.mp3	0:00 / 0:03	Вара күл буена.	2	1	thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bba6ea9918f2a9ad	tt_train_0/common_voice_tt_17343560.mp3	0:00 / 0:02	Матво түгел.	2	1	thirties	male_masculine		tt		

≡ tatar_cv.dict

1	-дип	d i p
2	ахмак	a x m a k
3	аз	a z
4	акчаның	a k ɕ a n ʏ ɲ
5	алты	a l t ʏ
6	авыртмас	a w ʏ r t m a s
7	Firefox	f a j r f o k s
8	IT	a j t i
9	аз	a z
10	ак	a k
11	азар	a z a r
12	ашасаң	a ʃ a s a ɲ


Alignment

- ★ Extract audio and transcriptions from the data and create matching .txt files for each .mp3 file

 common_voice_tt_17343437	MPEG Layer 3 Audio	59 KB
 common_voice_tt_17343437	Text Document	1 KB

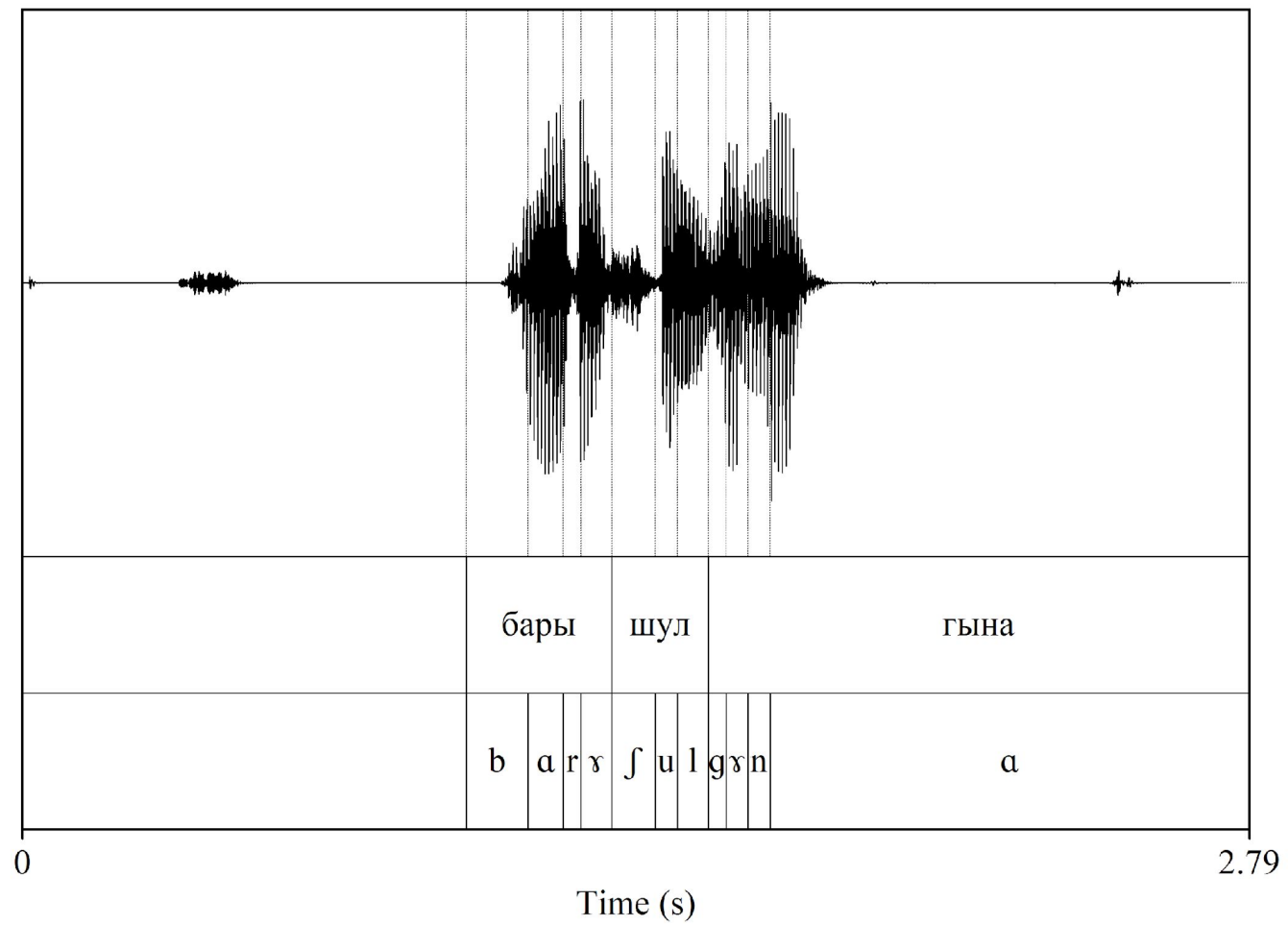
- ★ Run the aligner and export .TextGrid files for quality check

```
mfa align ~/data/raw tatar_model.zip tatar.dict ~/data/aligned
```

 common_voice_tt_17343437.TextGrid	TEXTGRID File	10 KB
---	---------------	-------

Analysis

- ★ Load samples of the generated files with their audio files in Praat
- ★ Check for deviations or general errors
- ★ So far: Even though the language is comparatively low-resource, the alignment works surprisingly well
- ★ However – the end of a word / sentence is often not found, last phoneme is tagged until the end of the recording
- ★ Some audio errors are present (chirping birds etc.), usually no influence



What remains to be done

- ★ Compare Tatar alignment quality to a high-resource language
 - Turkic (same language family, different script)
 - Russian (different language family, similar script and geographical region)
- ★ Compare different speakers / dialect regions
- ★ Create custom G2P dictionary
- ★ Use alignments to improve an existing ASR system

Resources

- ★ Montreal Forced Aligner <https://montreal-forced-aligner.readthedocs.io>
- ★ Praat <https://praat.org>
- ★ Common Voice Dataset <https://commonvoice.mozilla.org/de/datasets>

Thank you!
Any questions?