
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 rowsSplit (6)
train · 9.32k rows

Q Search this dataset

client_id string - classes	path string - lengths	audio audio - duration (s)	sentence string - lengths	up_votes int64	down_votes int64	age string - classes	gender string - classes	accent string - classes	locale string - classes	segment string - classes	variant string - classes
2 values	39 39	1.11 9.7	5 128	2	0	1 value	1 value	1 value	1 value	1 value	1 value
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343488.mp3	▶ 0:00 / 0:03	Баш есте, хан қаретларе.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343440.mp3	▶ 0:00 / 0:08	Тешенда биек коймалар белен өйләндәреп алингын бер бакчада бик...	3	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343441.mp3	▶ 0:00 / 0:06	Балык алеккен көнне тамаклары түк, балык алмакгөн көннө тамаклары ач...	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343445.mp3	▶ 0:00 / 0:02	Бу - минем фикер.	2	1	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343444.mp3	▶ 0:00 / 0:05	Ашап түйгөч, савытларыңда ашап батере алмаган сөтлөре кала.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343445.mp3	▶ 0:00 / 0:04	Татар түгел димә - башың өзтәрмән!	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343446.mp3	▶ 0:00 / 0:04	Бер көн бара, ике көн бара.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343447.mp3	▶ 0:00 / 0:03	Барысы да тынып калды.	3	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343449.mp3	▶ 0:00 / 0:03	Теге акданалы елгага кайтып жите.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343450.mp3	▶ 0:00 / 0:03	Бер алаша ала да чыгып ките.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343451.mp3	▶ 0:00 / 0:02	Барысы да кителэр.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343452.mp3	▶ 0:00 / 0:04	Кызык сезгө, яшь балалар, кызылы кичтө!	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343453.mp3	▶ 0:00 / 0:06	Сара Герей хан хатыннарын, балаларын, якнинарын алтын Волгар калкыснан...	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343556.mp3	▶ 0:00 / 0:04	Беренчедән, сез мина түгел, мин сезгө рахматле.	3	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343557.mp3	▶ 0:00 / 0:03	Бүгөн ял көне.	2	0	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad ec05d686304d568ab7063da28915d3fcfcf8d4dd4b81646bd13..	tt_train_0/common_voice_tt_17343559.mp3	▶ 0:00 / 0:03	Вара күл буена.	2	1	0 thirties	male_masculine		tt		
11d5e99f7bd5b4f8492a06bb1ec22aa9110bbadea9918f2a9ad	tt_train_0/common_voice_tt_17343560.mp3	▶ 0:00 / 0:02	Матын түгел...	2	1	0 thirties	male_masculine		tt		

< Previous 1 2 3 ... 94 Next >

≡ tatar_cv.dict

- 1 -дип d i p
- 2 ахмак a x m a k
- 3 аз a z
- 4 акчаның a k c a n u ɳ
- 5 алты a l t ɣ
- 6 авыртмас a w ɣ 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 ашасан a ſ a s a ɳ

Alignment

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

 common_voice_tt_17343437
 common_voice_tt_17343437

MPEG Layer 3 Audio 59 KB
Text Document 1 KB

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

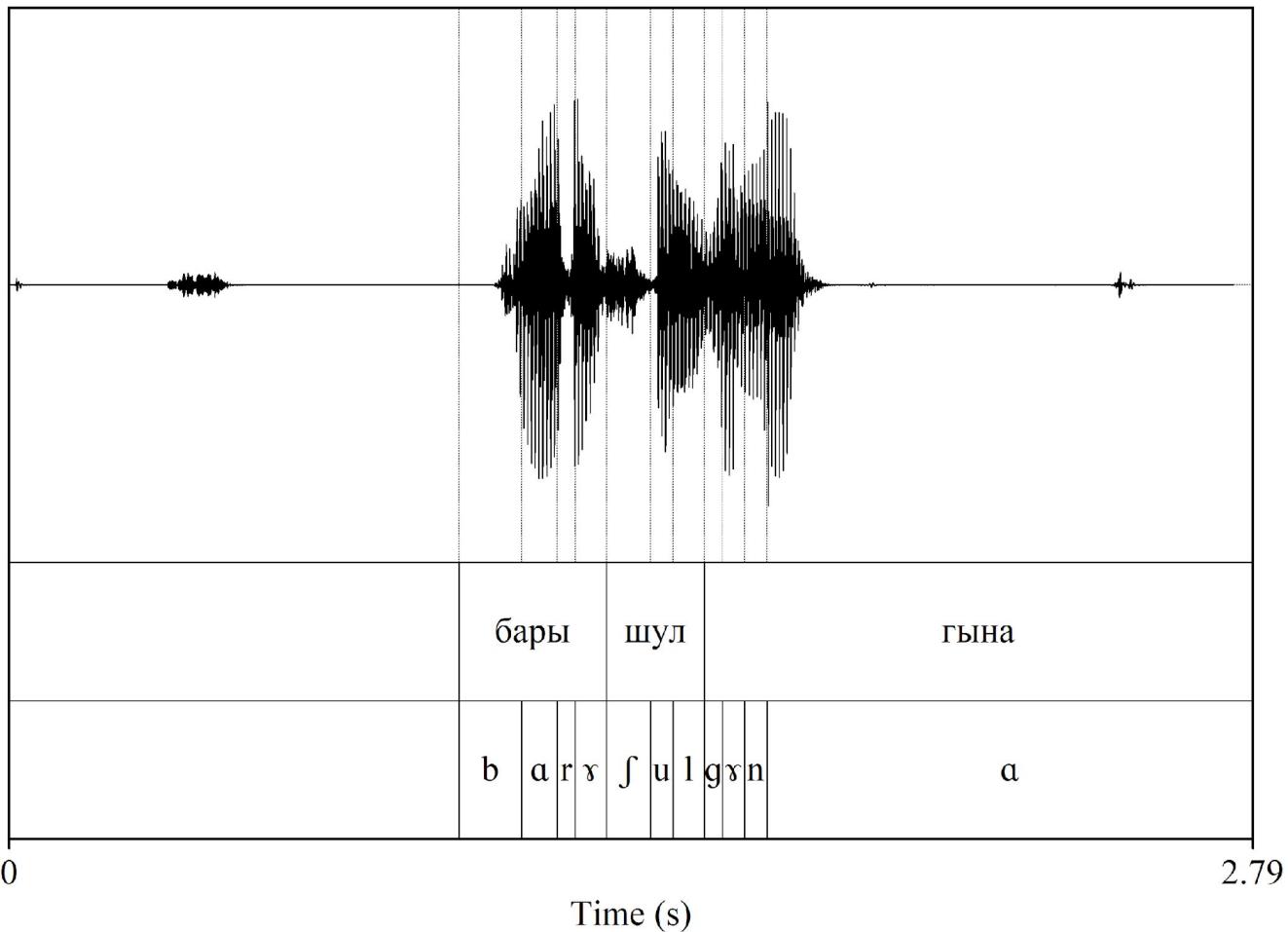
```
mfa align ~/data/raw tatar_model.zip tatar.dict ~/data/aligned
```

 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?