TTS and Data Selection: Improving Systems for Low-Resource Languages

Chevy Levitan, DREU 2015

outline

I. Project
II. Approach
III. Methods
IV. Status
V. Future

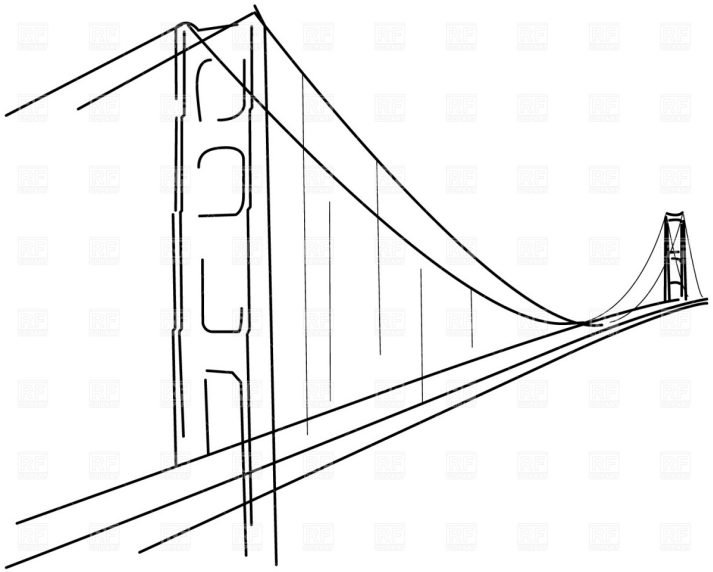
I.

Project

synthesize natural, intelligible voices for low resource languages using data selection

motivation

- bridge the gap



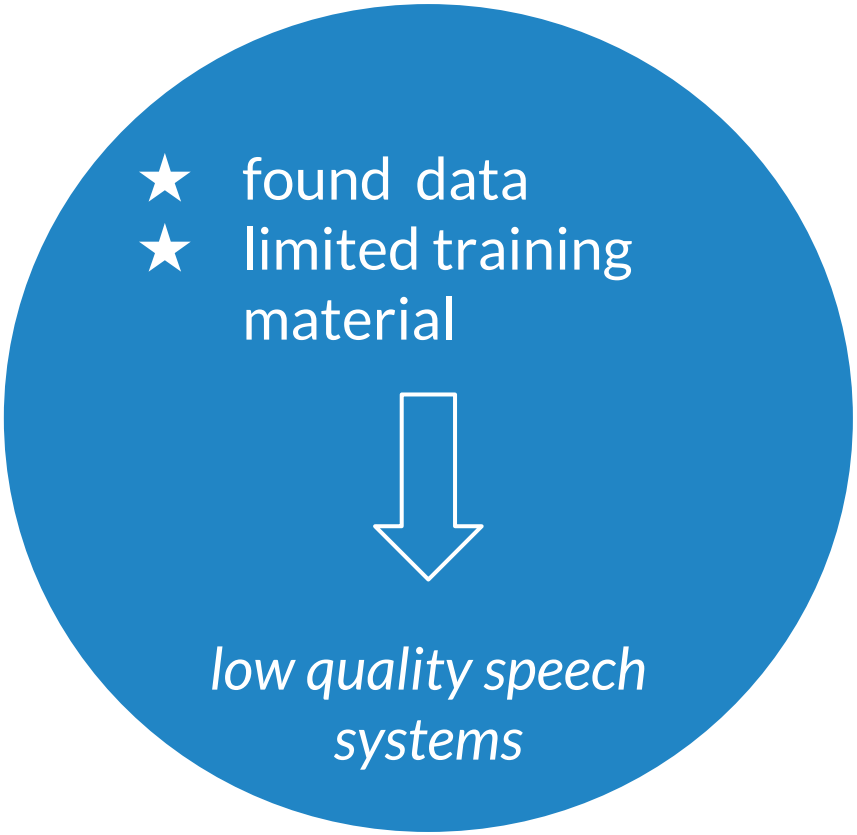
motivation

- bridge the gap
- allow for cross-language communication

why data selection?

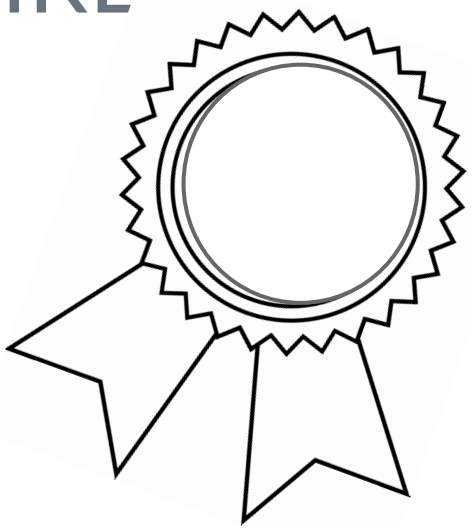
HRLs vs. LRLs


★ prepared data
★ abundance of training material
high quality speech systems


★ found data
★ limited training material
low quality speech systems

A. filter out unwanted data from training set

A. filter out unwanted data from training set
B. supplement limited LRL data with choice data from similar HRL



II.

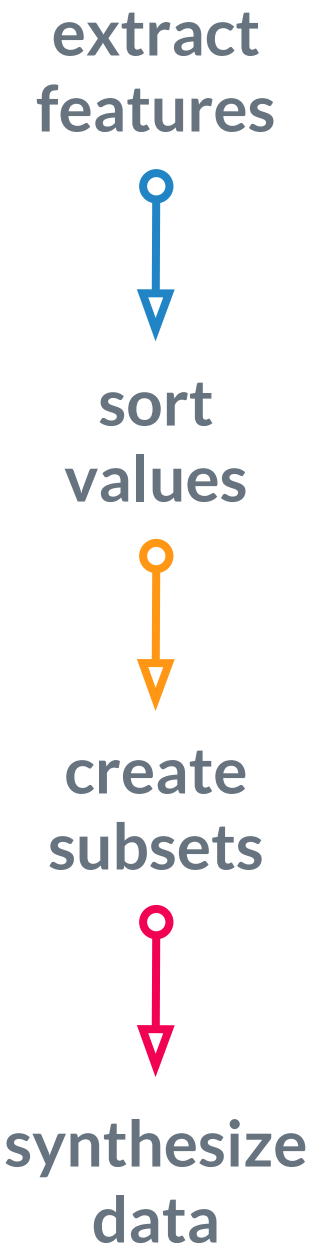
APPROACH

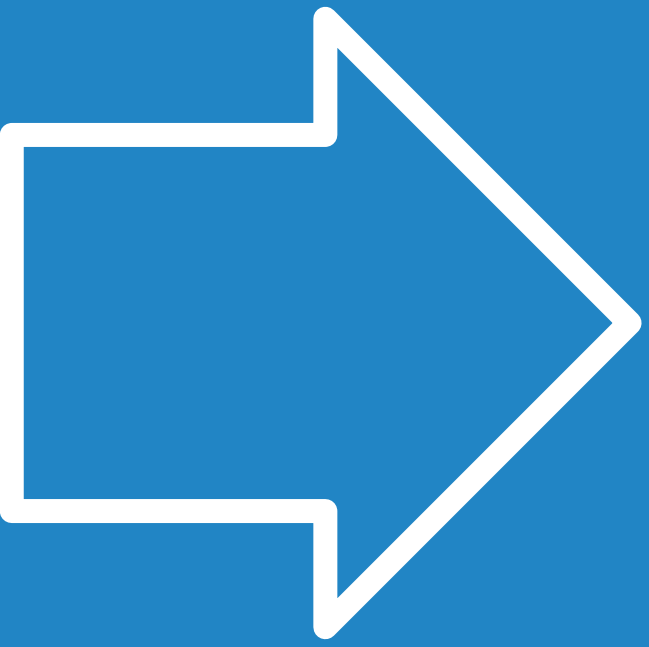
preparing the experiment

corpus

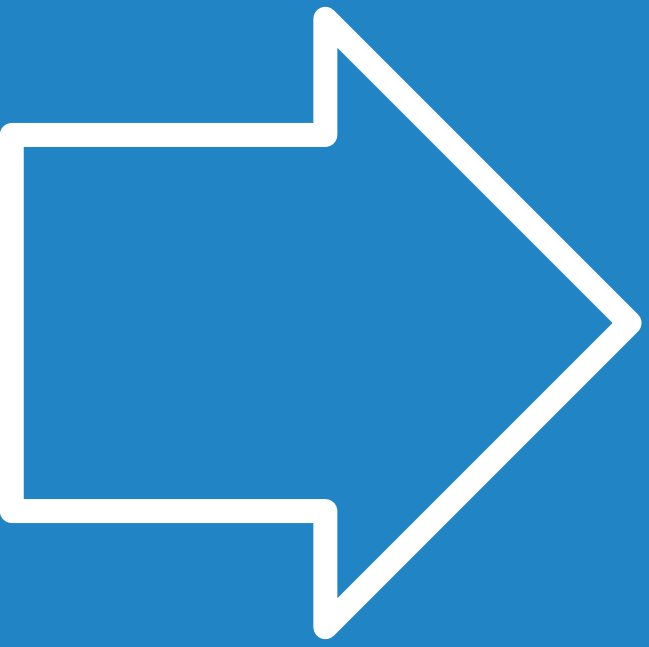
- Boston Radio News Corpus
- pre-processed
- English

data selection process





evaluate.



evaluate.

compare/contrast voices

example

VOICE 1

VOICE 2

solution

1. subset data
2. complete dataset

III.

METHODS

testing our hypothesis

standards

★ follow standard procedures for evaluating TTS voices

standards

- ★ follow standard procedures for evaluating TTS voices
- ★ successful voice = intelligible + natural

standards

- ★ follow standard procedures for evaluating TTS voices
- ★ successful voice = intelligible + natural
- ★ use crowdsourcing for unbiased results

mechanical turk

Intelligibility

- ➔ transcribe nonsense sentences
- ➔ accurate transcription = intelligible voice

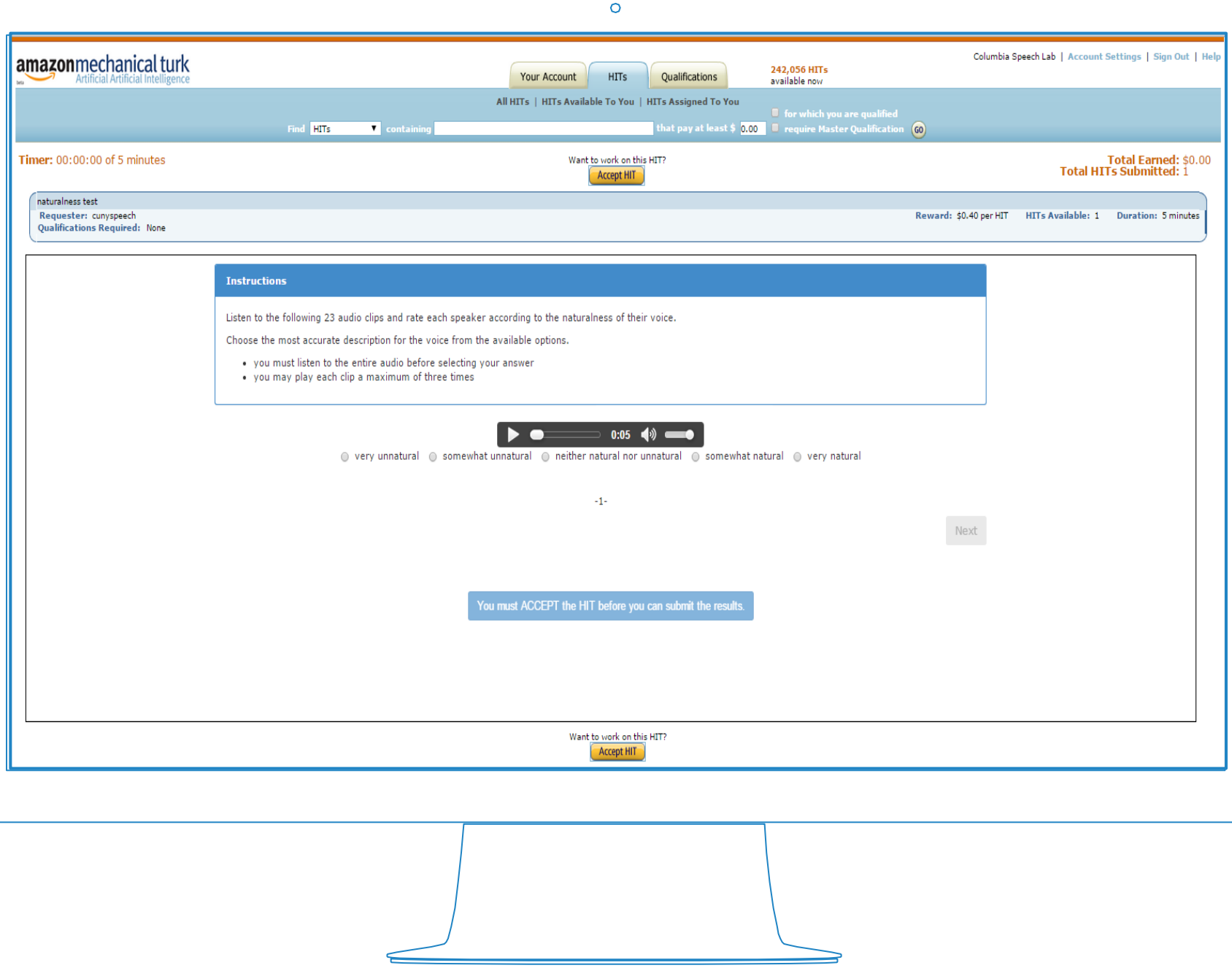
mechanical turk

Intelligibility

- ➔ transcribe nonsense sentences
- ➔ accurate transcription = intelligible voice

Naturalness

- ➔ use Likert scale to rate voices from very unnatural to very natural
- ➔ identify the voices are categorized as natural+


amazonmechanical turk
Artificial Artificial Intelligence
Columbia Speech Lab | Account Settings | Sign Out | Help
Your Account
HITs
Qualifications
242,056 HITs
available now
All HITs | HITs Available To You | HITs Assigned To You
Find HITs containing that pay at least $ 0.00 for which you are qualified require Master Qualification GO
Timer: 00:00:00 of 5 minutes
Want to work on this HIT?
Accept HIT
Total Earned: $0.00
Total HITs Submitted: 1
naturalness test
Requester: cunyspeech
Qualifications Required: None
Reward: $0.40 per HIT
HITs Available: 1
Duration: 5 minutes
Instructions
Listen to the following 23 audio clips and rate each speaker according to the naturalness of their voice.
Choose the most accurate description for the voice from the available options.
you must listen to the entire audio before selecting your answer
you may play each clip a maximum of three times
0:05
very unnatural
somewhat unnatural
neither natural nor unnatural
somewhat natural
very natural
-1-
Next
You must ACCEPT the HIT before you can submit the results.
Want to work on this HIT?
Accept HIT

IV.

STATUS

our current state

intelligibility HIT

✓ create subsets

intelligibility HIT

- ✓ create subsets
- ✓ synthesize voices with this data

intelligibility HIT

- ✓ create subsets
- ✓ synthesize voices with this data
- ✓ design and implement HIT

intelligibility HIT

- ✓ create subsets
- ✓ synthesize voices with this data
- ✓ design and implement HIT
- ✓ publish on MTurk site

intelligibility HIT

- ✓ create subsets
- ✓ synthesize voices with this data
- ✓ design and implement HIT
- ✓ publish on MTurk site
- ✓ workers complete HITs

intelligibility HIT

- ✓ created subsets
- ✓ synthesized voices with this data
- ✓ design and implement HIT
- ✓ publish on MTurk site
- ✓ workers complete HITs
- ✓ accept/reject work

naturalness HIT

- ✓ create subsets

naturalness HIT

- ✓ create subsets
- ✓ synthesize voices with this data

naturalness HIT

- ✓ create subsets
- ✓ synthesize voices with this data
- ✓ design and implement HIT

naturalness HIT

- ✓ create subsets
- ✓ synthesize voices with this data
- ✓ design and implement HIT
  - publish on MTurk site
  - workers complete HITs
  - accept/reject work

V.

FUTURE

further exploration of this research

evaluation

analyze mechanical turk responses

evaluation

analyze mechanical turk responses

low-resource

implement data selection for LRLs

evaluation

analyze mechanical turk responses

low-resource

implement data selection for LRLs

text

apply similar methods to automatically select text data

Thanks!

Any questions?