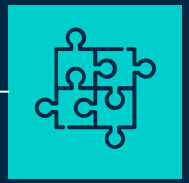
ANALYZING BIAS IN AUTOMATED TRANSCRIPTION SERVICES FOR ACCENTED SPEECH

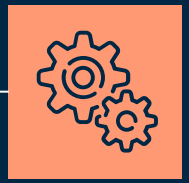
Aastha Goyal, Rupali Roy &
Shravya Kolavara

AGENDA



01

PROBLEM STATEMENT & RELATED WORK

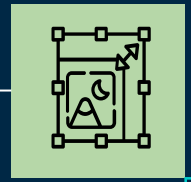


02

DATA SET DESCRIPTION & METHODS

03

ANALYSIS & TAKEAWAYS

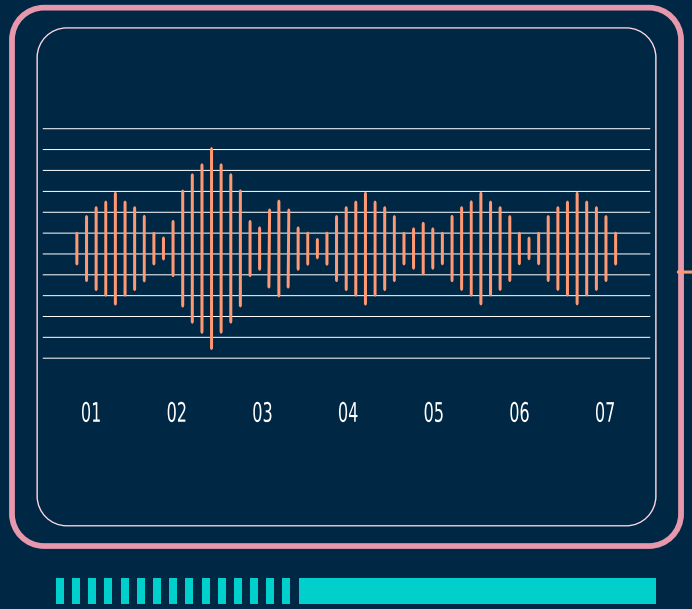


04

PROPOSED DESIGN & CHALLENGES

PROBLEM DEFINITION

Bias present in current state of the art transcription softwares caused by racial and demographic disparities in accent which leads to lower accuracy of transcribed files for non-native English speakers.



RELATED WORK

Lack of Data

Disparities with regards to performance gap for black speakers as compared to white speakers.

Biased Training data

Due to phonetic mismatch between native and non-native english speakers, the models struggled.

Disparities cause harm

Speech recognition used by employers, criminal justice agencies and educational platforms.

Project overview – What is our goal?

To first identify the bias against accented speech in commercial ASR systems and then to help mitigate that bias by designing a crowdsourced database used for training models.

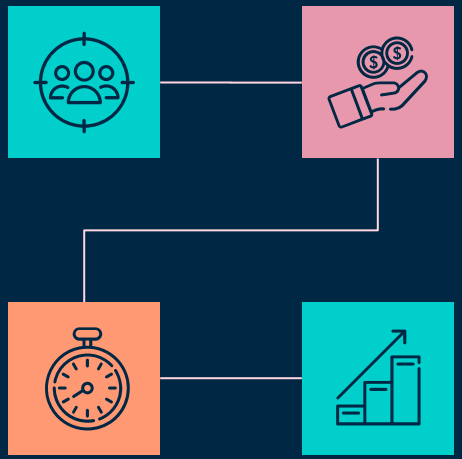
METHOD

Collecting Data

Generating Transcription of speech samples on various portals

Accuracy

Finding the Word Error Rate (WER) for all speech sample transcriptions



Transforming Data

Cleaning and mapping transcriptions to user details

Analysis & Design

Analysing bias due to native language and age onset and designing a solution

Data Set Description

2172	214	168	75
Voice Samples	Native Languages	Countries	Countries with less than 5 samples

Passage Text:
Please call Stella. Ask her to bring these things with her from the store: Six spoons of fresh snow peas, five thick slabs of blue cheese, and maybe a snack for her brother Bob. We also need a small plastic snake and a big toy frog for the kids. She can scoop these things into three red bags, and we will go meet her Wednesday at the train station.

Weinberger, Steven. (2015). *Speech Accent Archive*. George Mason University. Retrieved from http://accent.gmu.edu

Analysis of Passage Text

The paragraph uses common English words, but contains a variety of difficult English sounds and sound sequences.

The paragraph contains practically all of the sounds of English.

The Sounds in the Elicitation Paragraph (numbers indicate occurrences)

single consonants

initial		final	
k	(3)	z	(5)
t	(3)	l	(4)
ð	(6)	ŋ	(1)
θ	(3)	θ	(1)
w	(5)	m	(1)
s	(2)	ɹ	(5)
f	(3)	v	(3)
tʃ	(1)	ʃ	(1)
n	(1)	k	(4)
b	(3)	b	(1)
l	(1)	d	(2)
ʃ	(2)	g	(2)
d	(1)	n	(4)
ɹ	(1)	p	(1)
g	(1)	t	(2)
m	(2)		
h	(4)		

vowels

i	(12)
ɑ	(4)
ɛ	(4)
æ	(10)
ɪ	(11)
ʌ	(2)
ə	(10)
u	(5)
oʊ	(3)
aɪ	(1)
eɪ	(5)
ɔ	(3)
ɔɪ	(1)

clusters

initial		final	
pl	(2)	sk	(1)
st	(4)	ŋz	(2)
bɹ	(2)	ks	(1)
fɹ	(3)	nz	(2)
sp	(1)	bz	(1)
sn	(3)	nd	(3)
sl	(1)	dz	(1)
bl	(1)	gz	(1)
sm	(1)		
sk	(1)		
θɹ	(1)		
tɹ	(1)		

Survey Questions vs Available Data Variables

1. Where were you born?
2. What is your native language?
3. What other languages besides English and your native language do you know?
4. How old are you?
5. How old were you when you first began to study English?
6. How did you learn English? (academically or naturalistically)
7. How long have you lived in an english-speaking country? Which country?
8. What is your gender?

1. Age
2. Age Onset
3. Birthplace
4. Native Language
5. Birth Country
6. Sex

Top 10 sampled Native Languages

Native Language	Sample Count
English	579
Spanish	162
Arabic	102
Mandarin	65
French	63
Korean	52
Russian	48
Portuguese	48
Dutch	47
Turkish	37

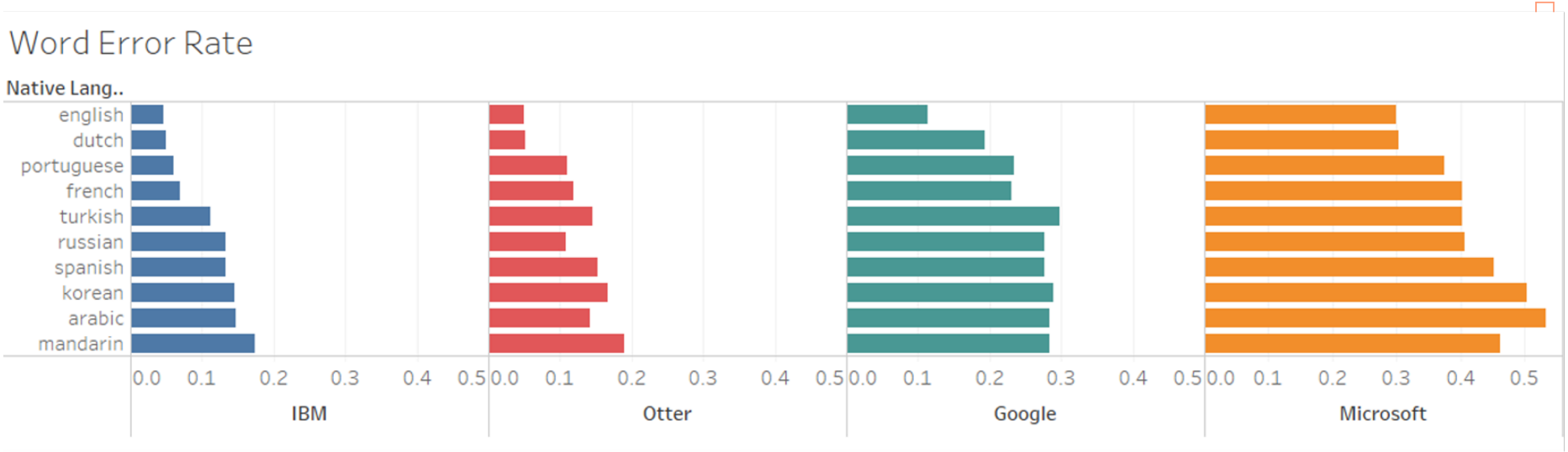
Analysis

Accuracy Metric:

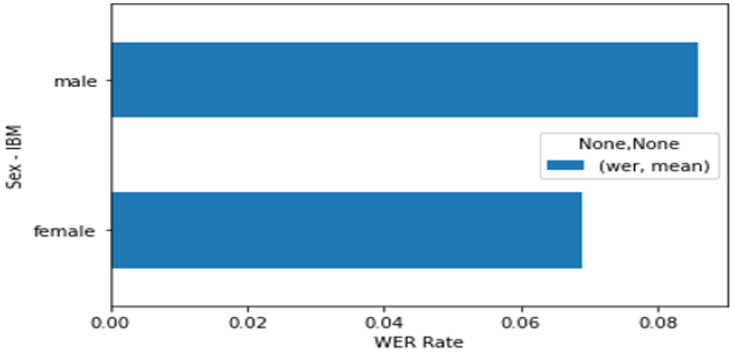
Word Error Rate (WER):
WER is defined as number of errors divided by the total words. We have used Jiwer Python Package to get this.

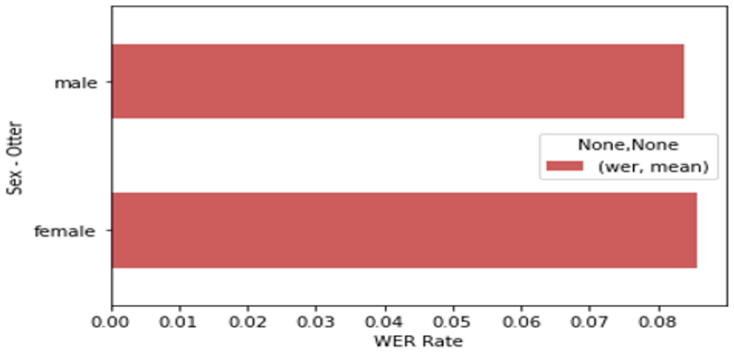
Compare WER for these four commercial state-of-art voice transcription systems

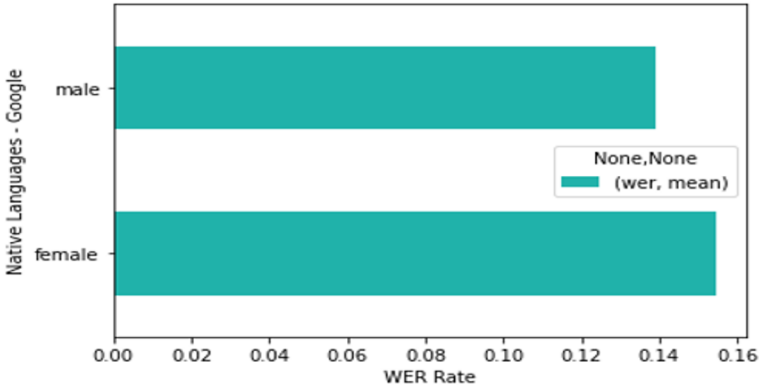
Word Error Rate – Top 10 Native Language

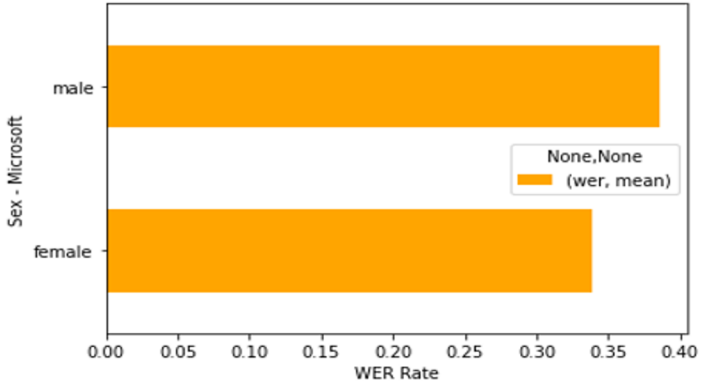


Word Error Rate vs Gender

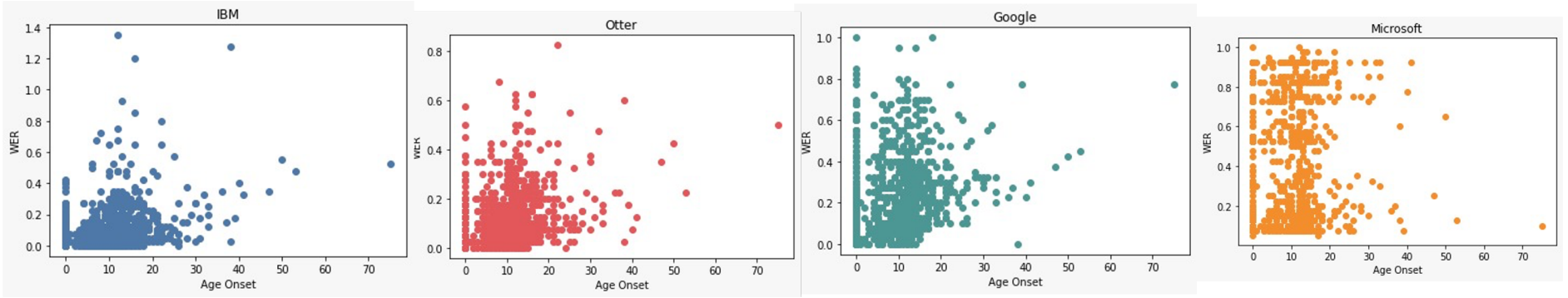








Word Error Rate – Age Onset



Takeaways

- Limited speech samples for non-native languages.
- There is a bias that exists against non-native English speakers when transcribing data using these commercial ASR systems.
- Systems work better for people who start learning English at an earlier age.
- The current dataset only considers male and female speakers, we can expand it to inlcude samples from transgender.

How can we collect diverse data in order to accommodate a diverse set of english speakers??

CREATE A OPEN SOURCE, CROWDSOURCED DATASET OF ACCENTED SPEECH.

Speech Accent Archive Analysis

Purpose: “To compare the demographic and linguistic backgrounds of the speakers in order to determine which variables are key predictors of each accent. The speech accent archive demonstrates that accents are systematic rather than merely mistaken speech.”

Pros: Clear objective, instructions to submit a sample, native phonetic inventory, and browse by language or atlas.

Cons: Only one elicitation paragraph, experimental environment, only researchers can submit samples, missing data and no interactive elements.

EXISTING DATABASES

Databases for specific ASR tasks

AUDIOSET

VOXCELEB

Accent related Databases

FOREIGN ACCENTED ENGLISH (FAE) CORPUS

CMU-ARCTIC

COMMON VOICE

Proposed Database

CROWDSOURCED

Intrinsic motivation and personalization can be used as motivating factors.

INTERACTIVE

Interactive maps and graphs provides users easily understand the data that is represented.

TRANSPARENT

It is highly important for the dataset to be transparent about the how, what and why of the data being collected.

GAMIFIED

This will attract more users to “donate their voice” to the dataset.

Our Website

https://rupaliroy30.wixsite.com/haiproject

Q&A