

# DIALECT DETECT

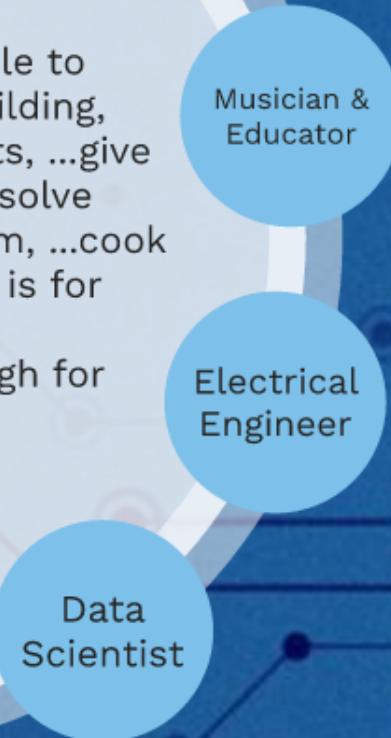
Steven Becerra



# Me

"A human being should be able to change a diaper, ...design a building, write a sonnet, balance accounts, ...give orders, cooperate, act alone, solve equations, analyze a new problem, ...cook a tasty meal, ... Specialization is for insects."

— Robert Heinlein, Time Enough for Love



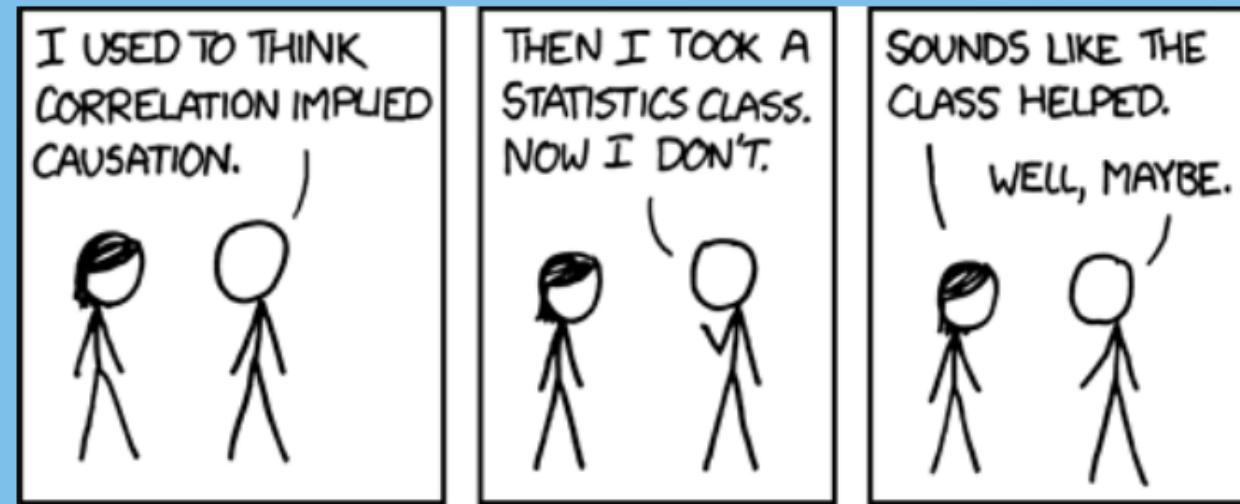
# Musician & Educator



# Electrical Engineer



# Data Scientist galvanize



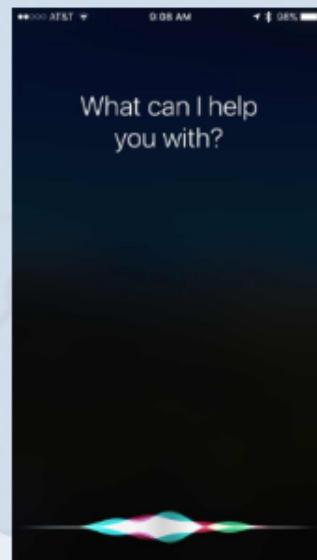
# DIALECT DETECT

Steven Becerra



# The Problem

Can I predict your native language?



The  
Data

The  
Phrase

# The Speech Accent Archive

Almost 2500 audio samples from over 300 countries



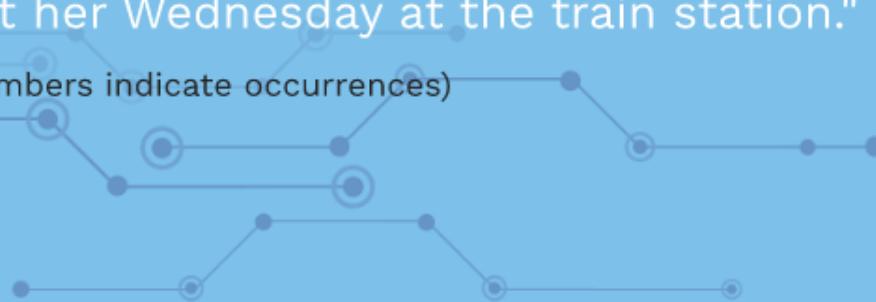
Weinberger, Steven. (2015). Speech Accent Archive. George Mason University. Retrieved from <http://accent.gmu.edu>

# The Phrase

"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."

The Sounds in the Elicitation Paragraph (numbers indicate occurrences)

single consonants	final	vowels	initial	clusters	final
k (3)	z (5)	i (12)	pl (2)	sk (1)	
t (3)	l (4)	a (4)	st (4)	nz (2)	
d (6)	r (1)	e (4)	br (2)	ks (1)	
θ (3)	θ (1)	æ (10)	tr (3)	nz (2)	
w (5)	m (1)	ɪ (11)	sp (1)	bz (1)	
s (2)	ɹ (5)	ʌ (2)	sn (3)	nd (3)	
f (3)	v (3)	ə (10)	sl (1)	dz (1)	
tʃ (1)	ʃ (1)	ʊ (5)	bl (1)	gz (1)	
n (1)	k (4)	ɔə (3)	sm (1)		
b (3)	b (1)	aɪ (1)	sk (1)		
l (1)	d (2)	ər (5)	θr (1)		
j (2)	g (2)	ɔ (3)	tr (1)		
d (1)	n (4)	ər (1)			
ɹ (1)	p (1)				
g (1)	t (2)				
m (2)					
h (4)					



# DIALECT DETECT

Steven Becerra



## The How

Audio

MFCC

Neural  
Network

# Audio

3 Languages

English  
Arabic  
Mandarin

73 Audio  
Samples  
Per  
Language



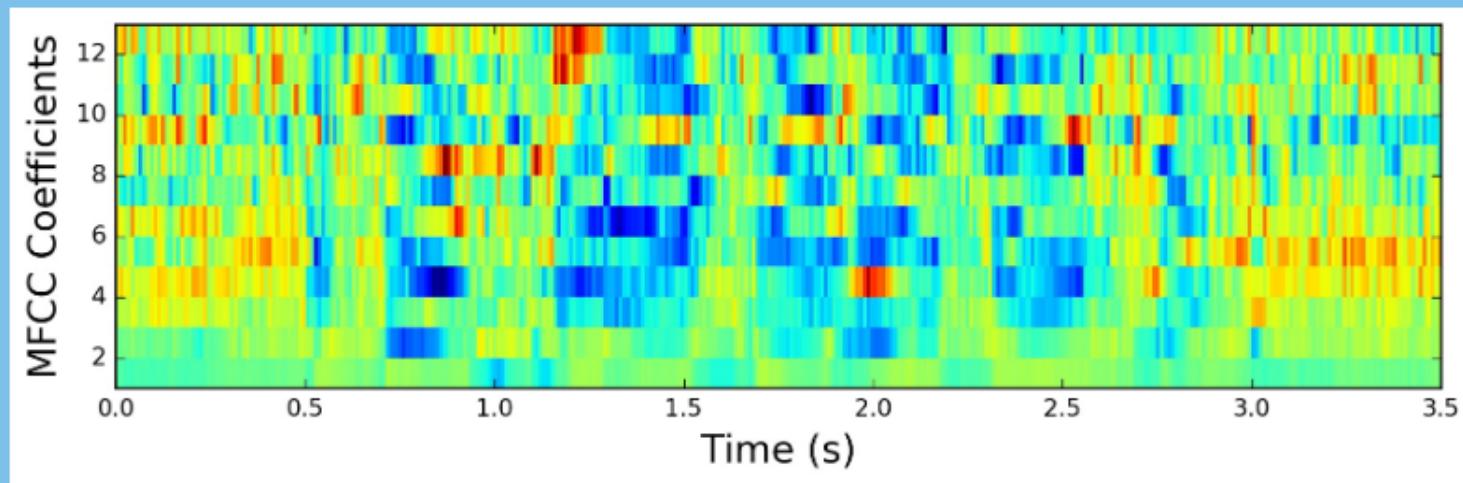
# 3 Languages

English  
Arabic  
Mandarin

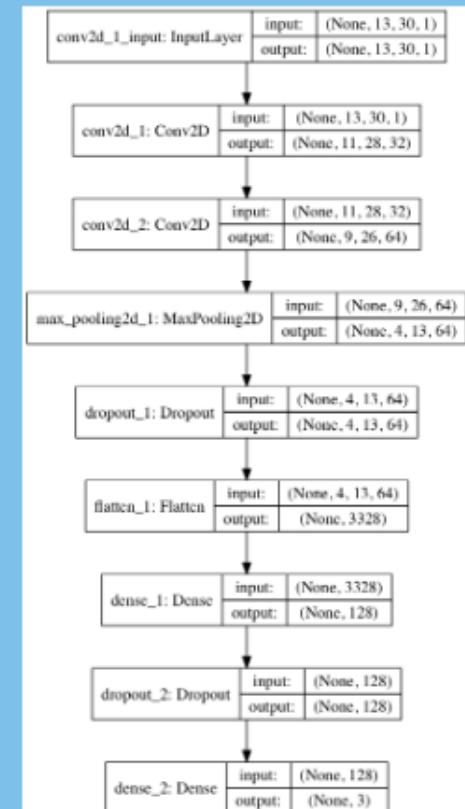
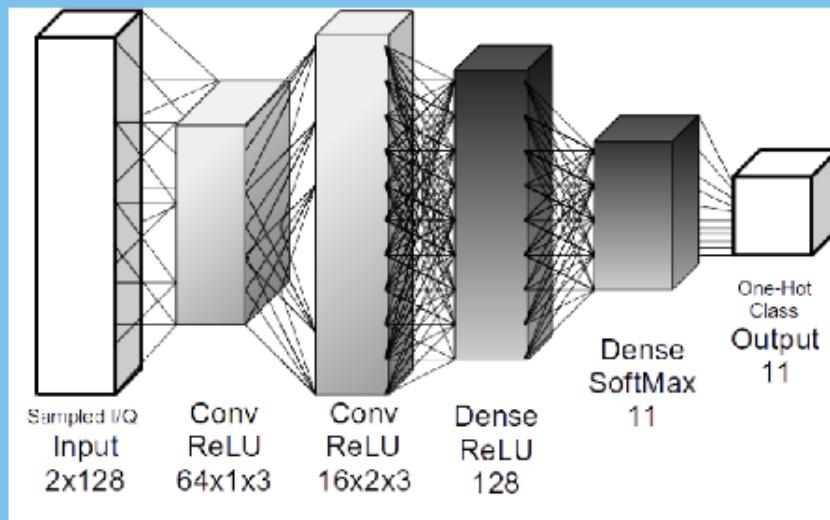
73 Audio  
Samples  
Per  
Language



# Mel Frequency Cepstral Coefficients



# 2D Convolutional Neural Network



# DIALECT DETECT

Steven Becerra



# Challenges & Solutions

Computationally  
Intensive

Small  
sample  
size

Still a  
small  
sample  
size

Results

# Computationally Intensive



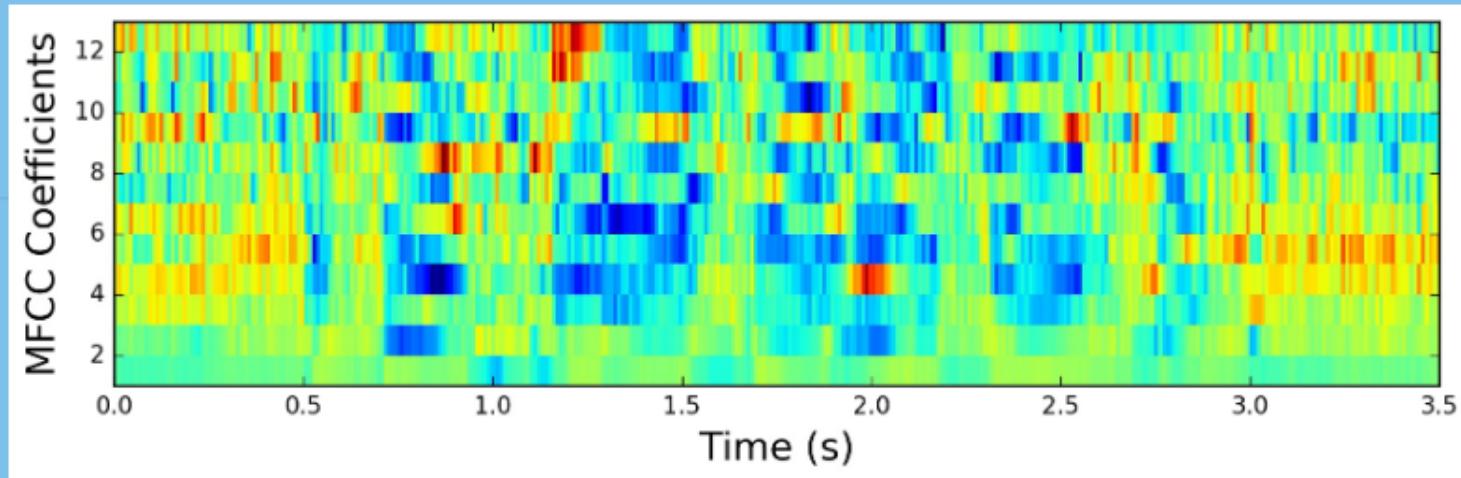
94 EC2  
Compute Units  
(32 virtual  
cores)

8 NVIDIA K80  
(GK210) GPUs

Python's  
multiprocessing  
package to handle  
audio processing

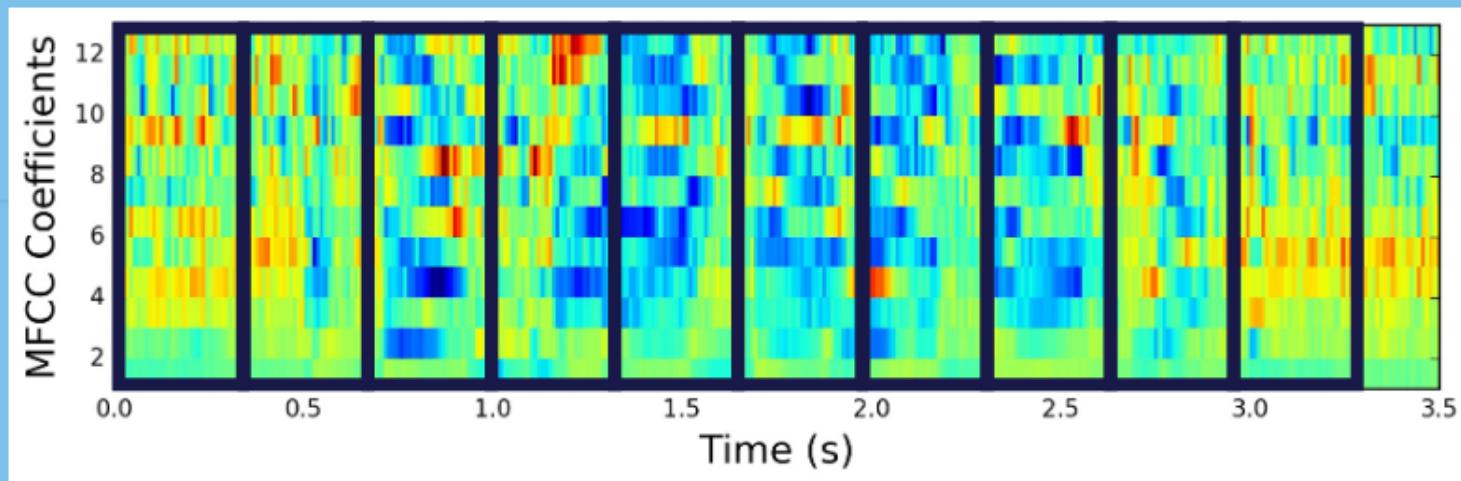
# Small sample size

Slicing & Ensembling



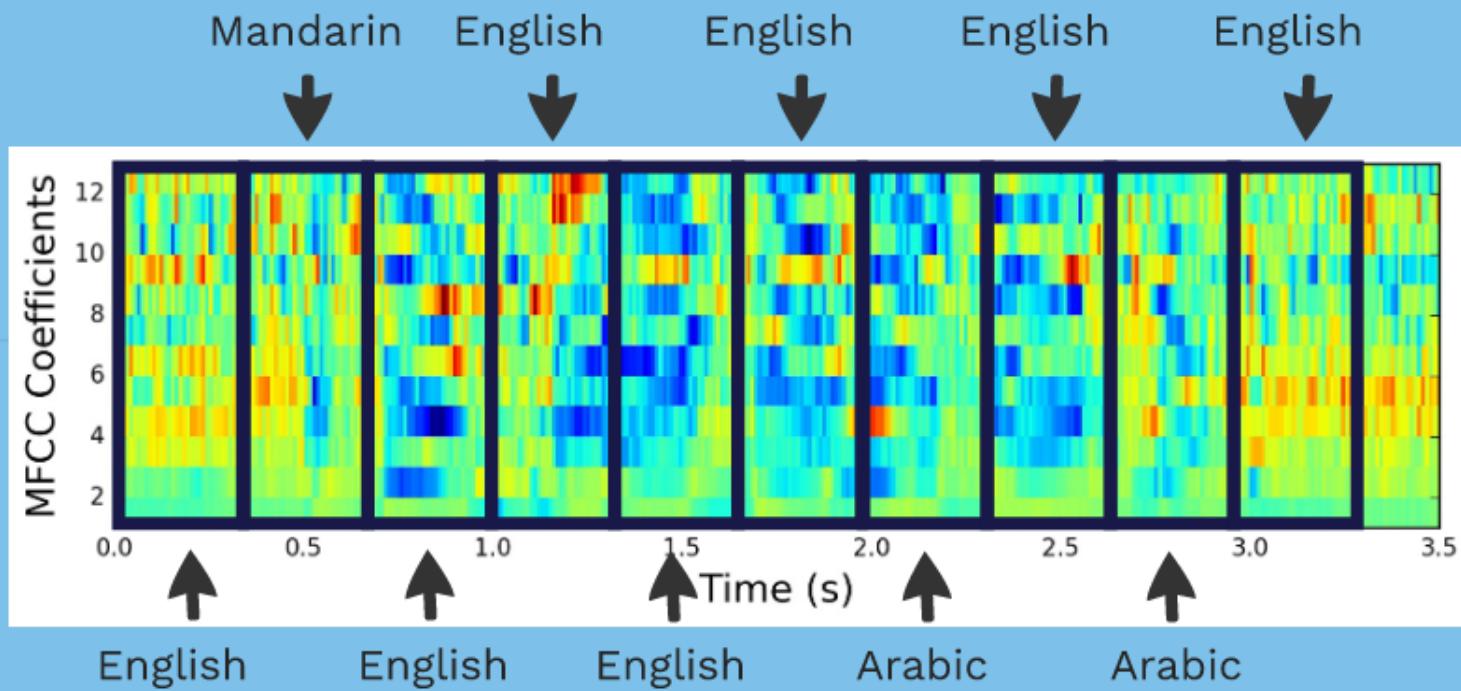
# Small sample size

Slicing & Ensembling



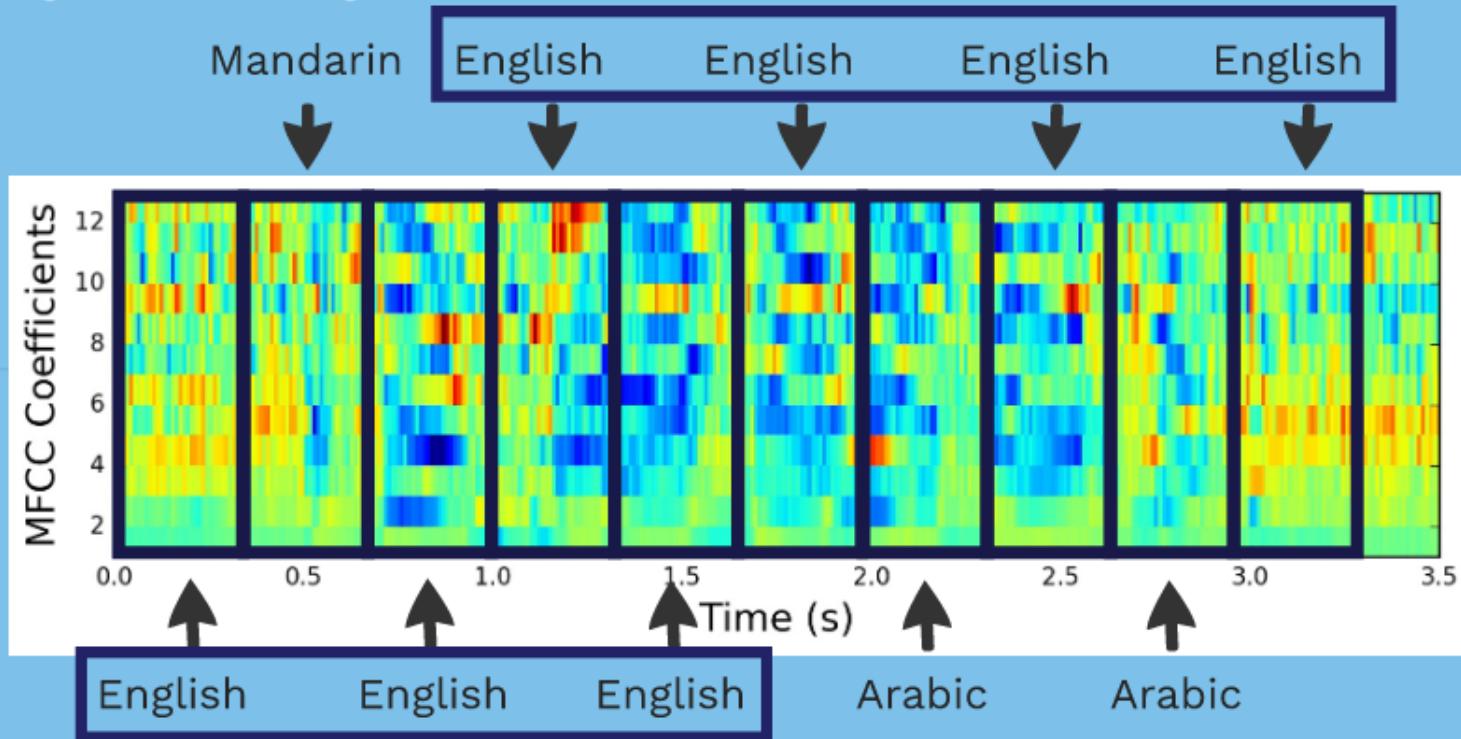
# Small sample size

Slicing & Ensembling



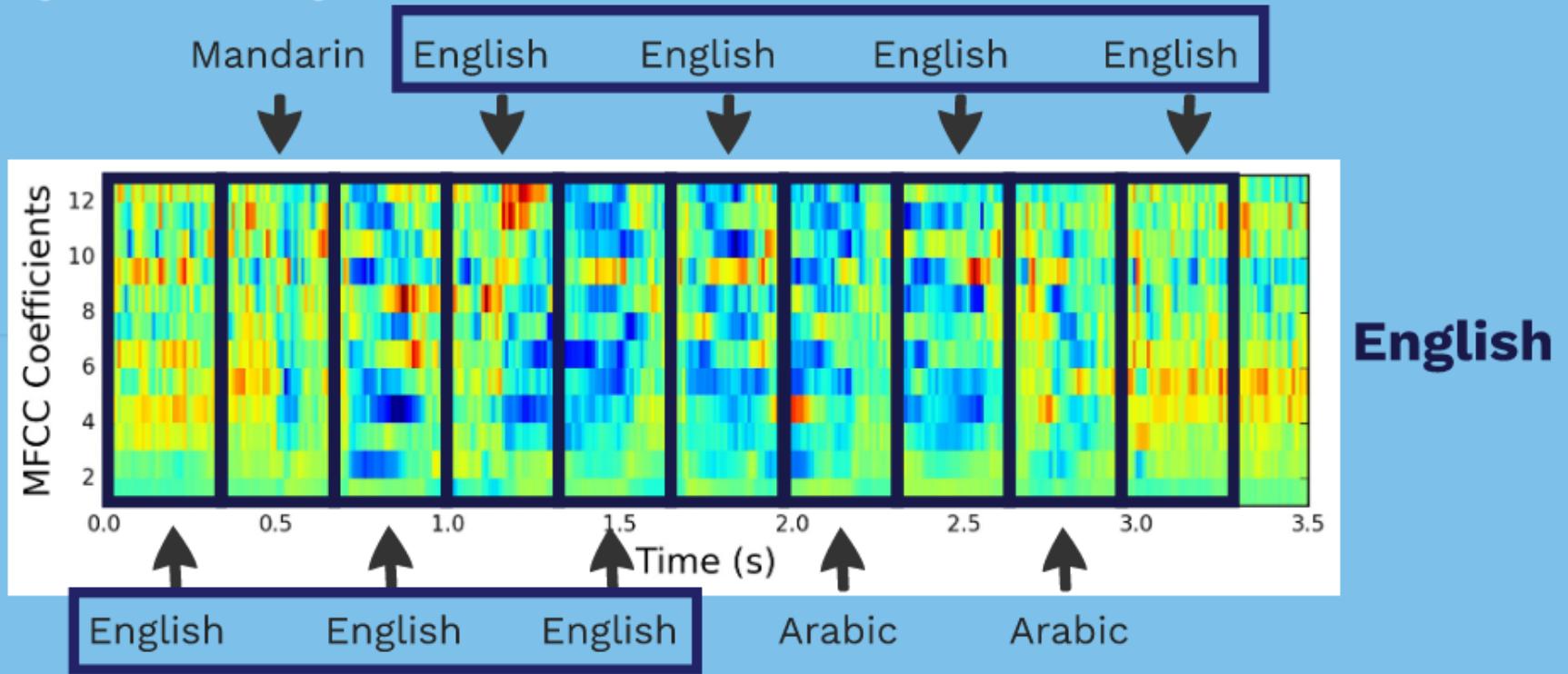
# Small sample size

Slicing & Ensembling



# Small sample size

Slicing & Ensembling



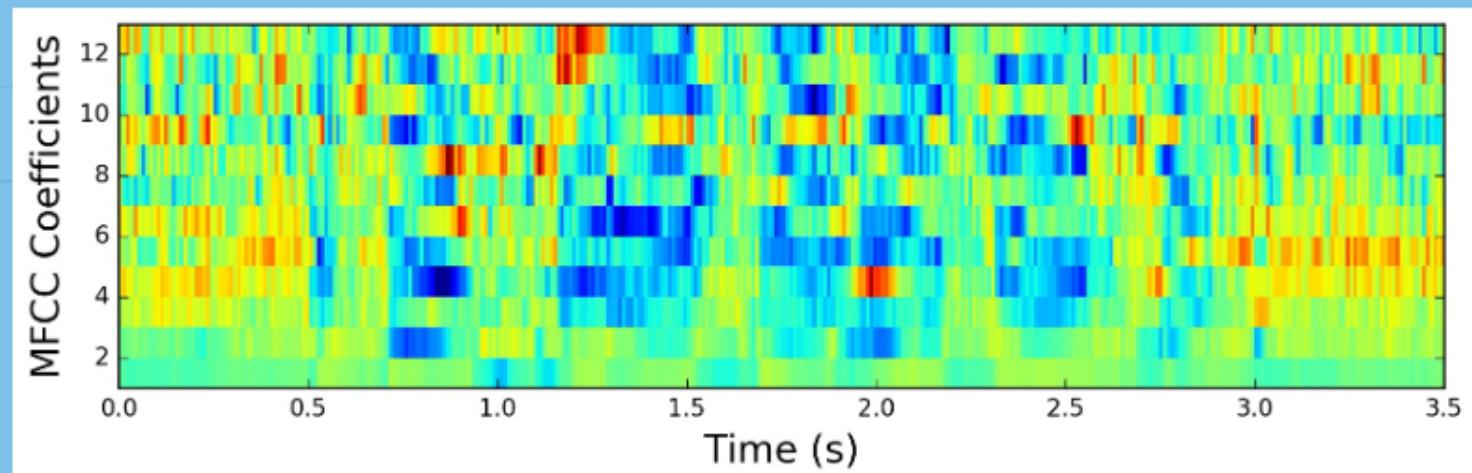
# Still a small sample size

Horizontal Image Shifting



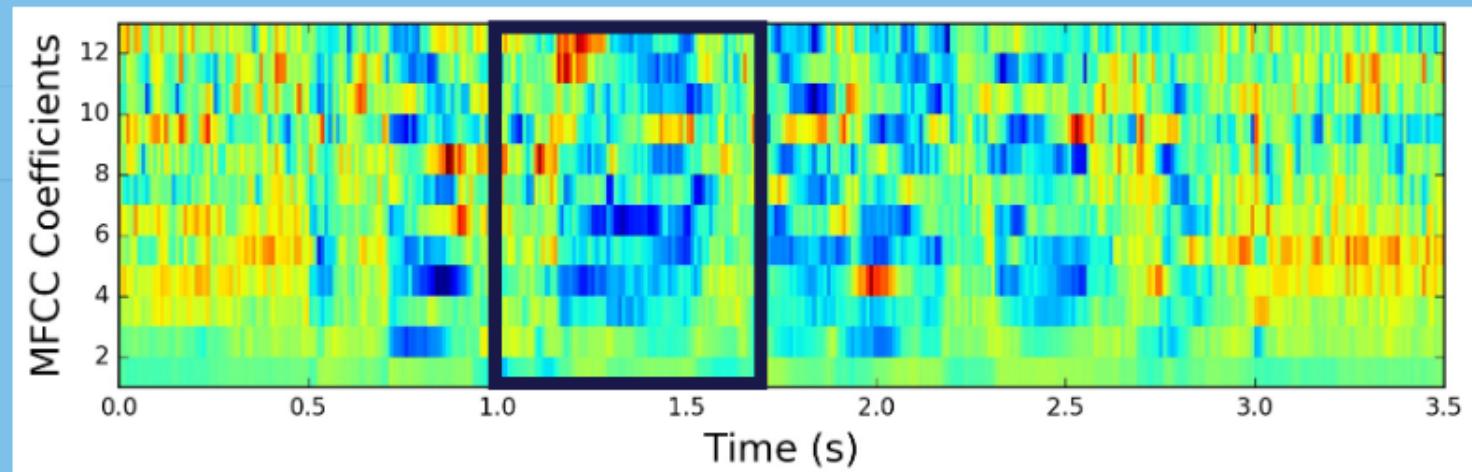
# Still a small sample size

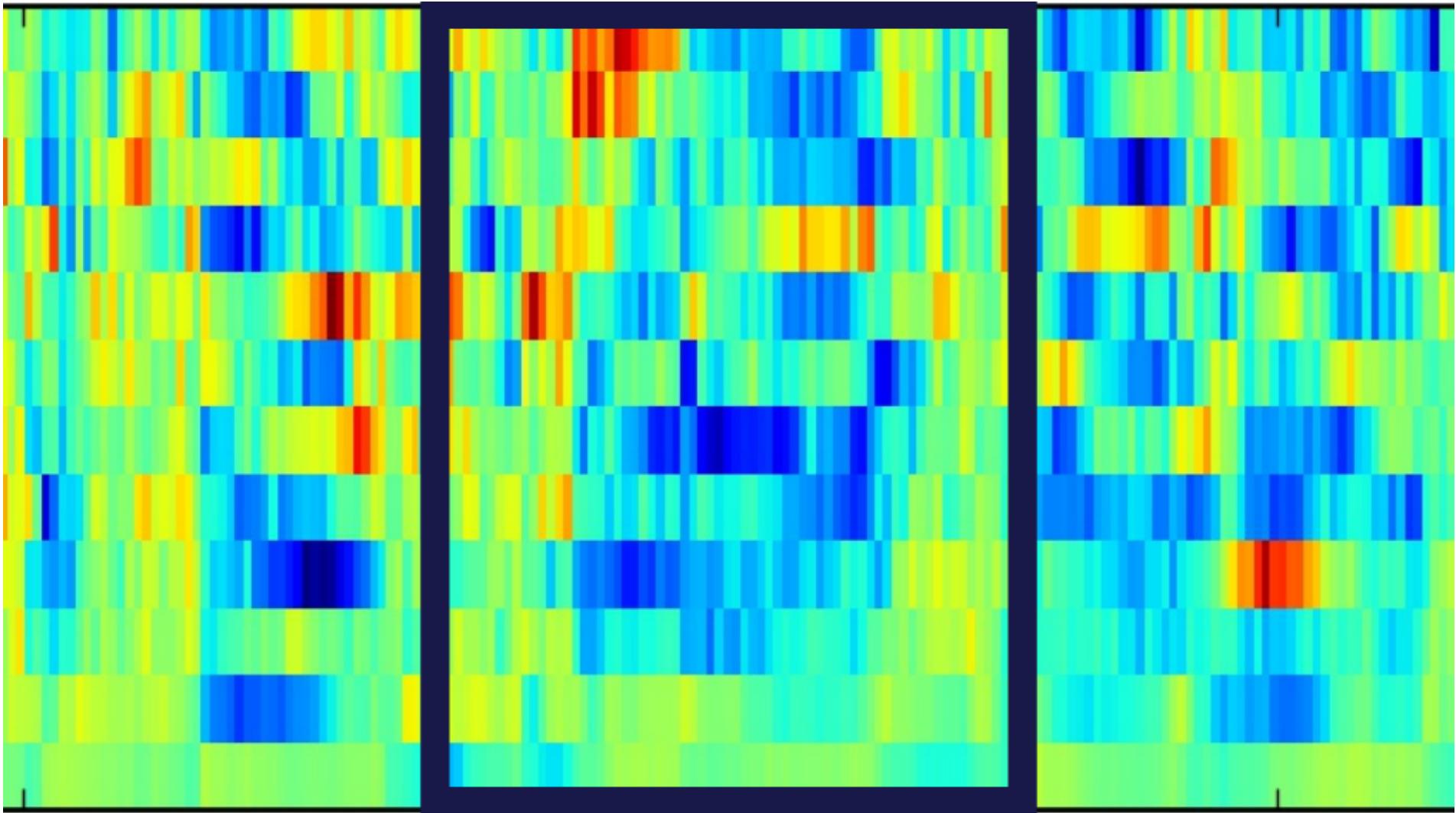
Horizontal Image Shifting



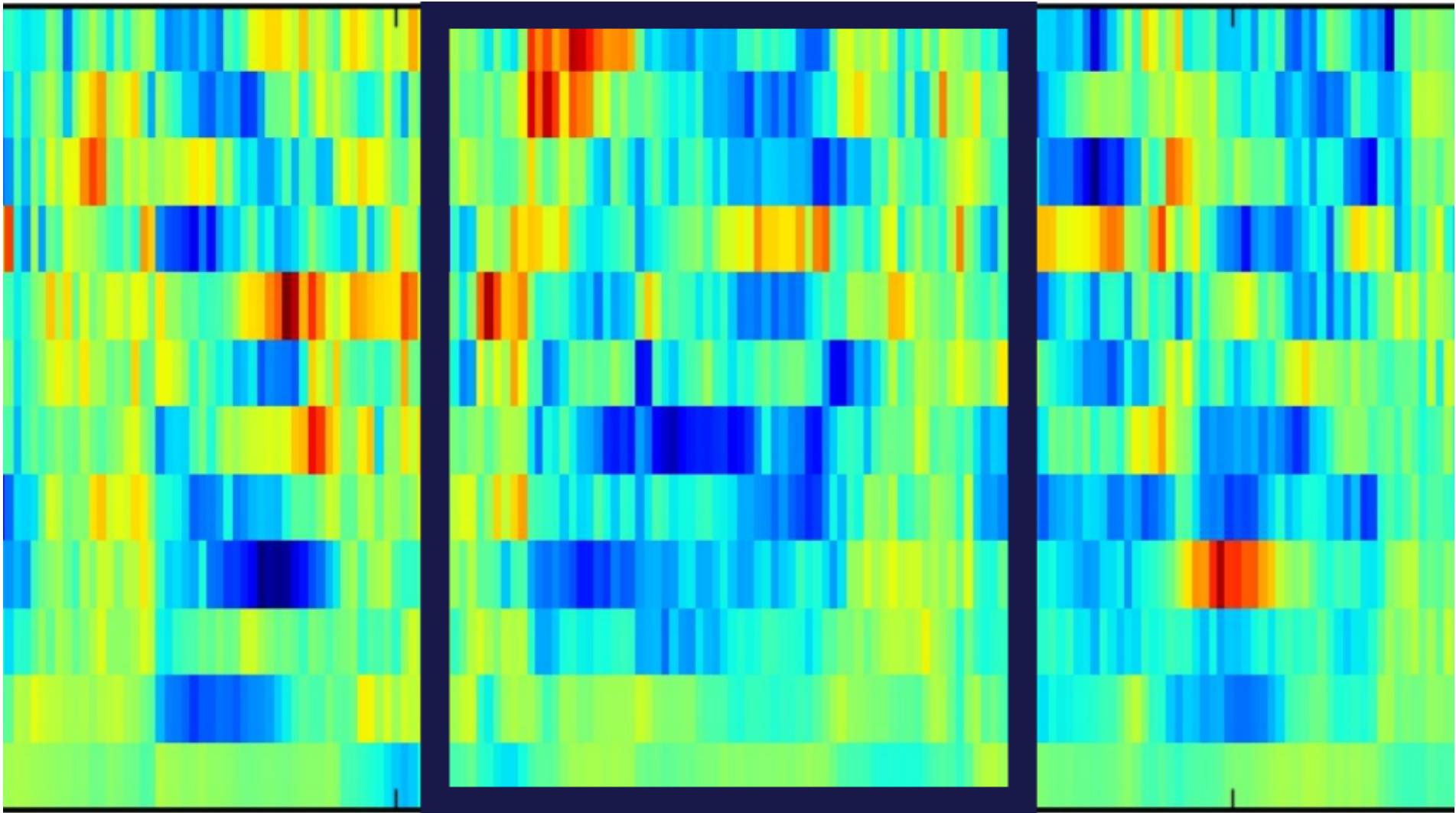
# Still a small sample size

Horizontal Image Shifting

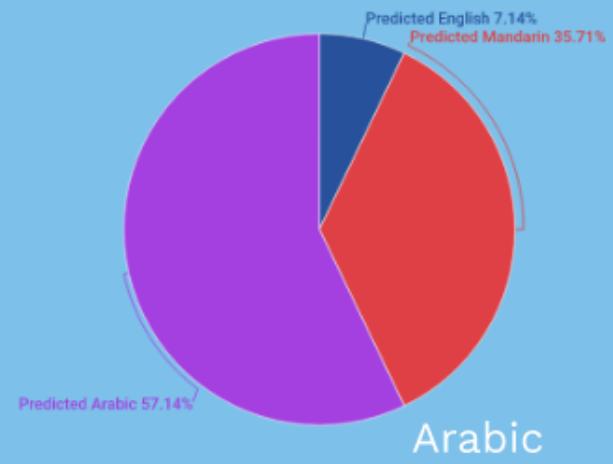
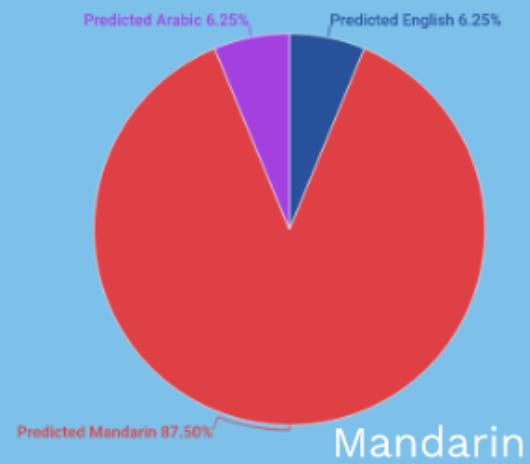
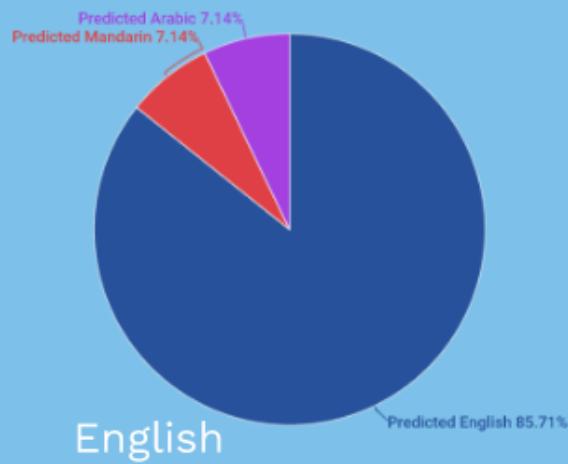


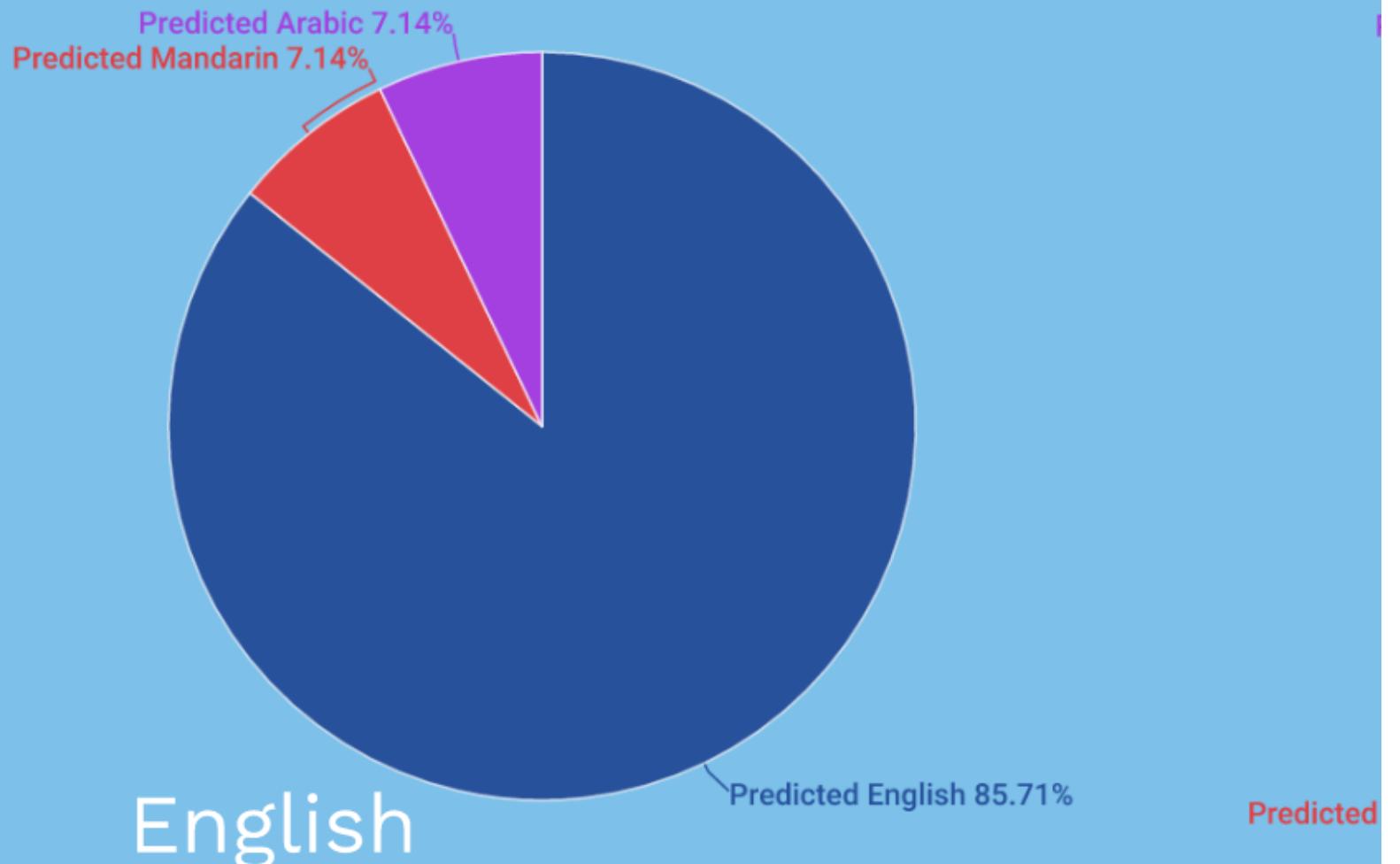


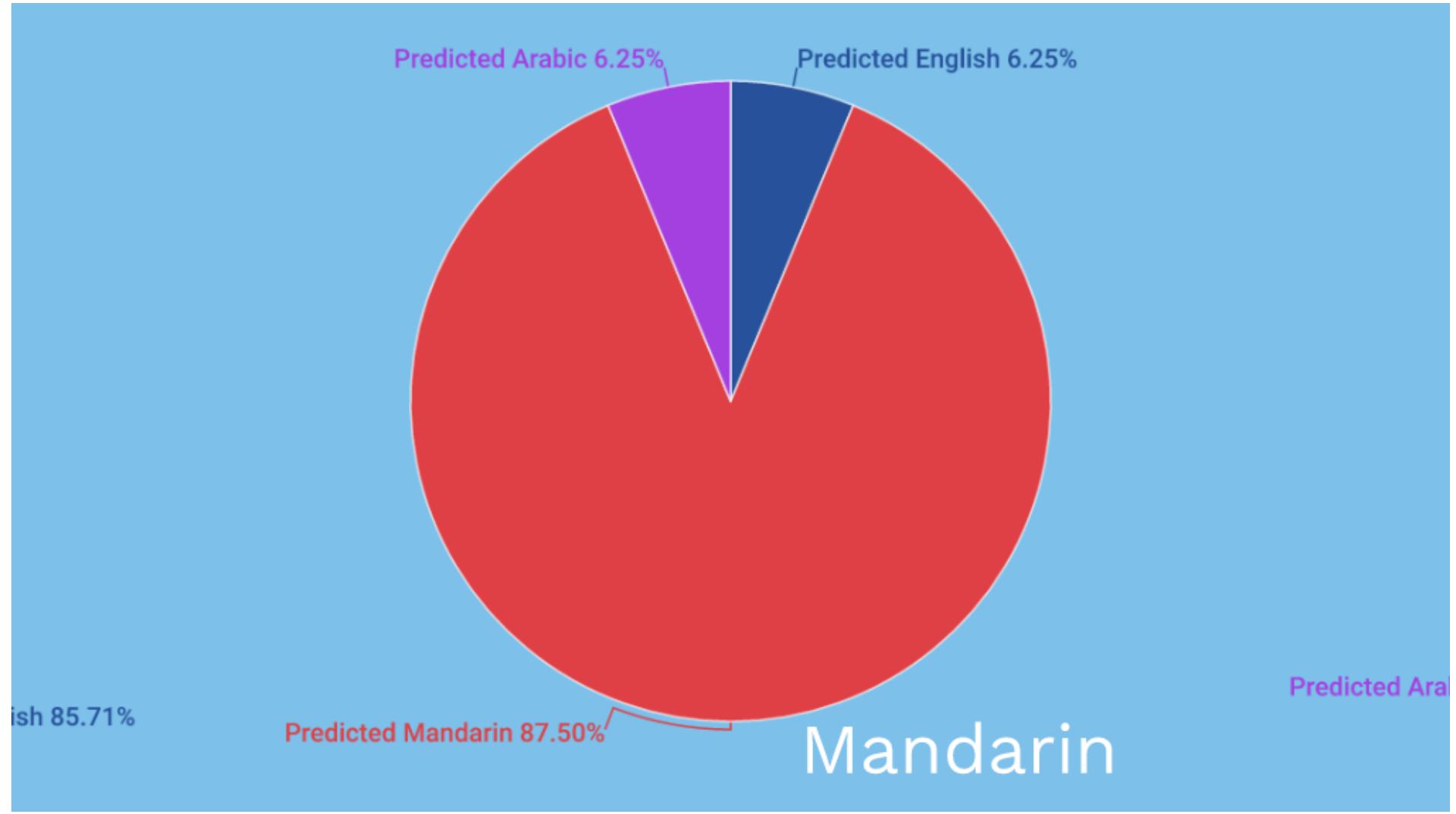


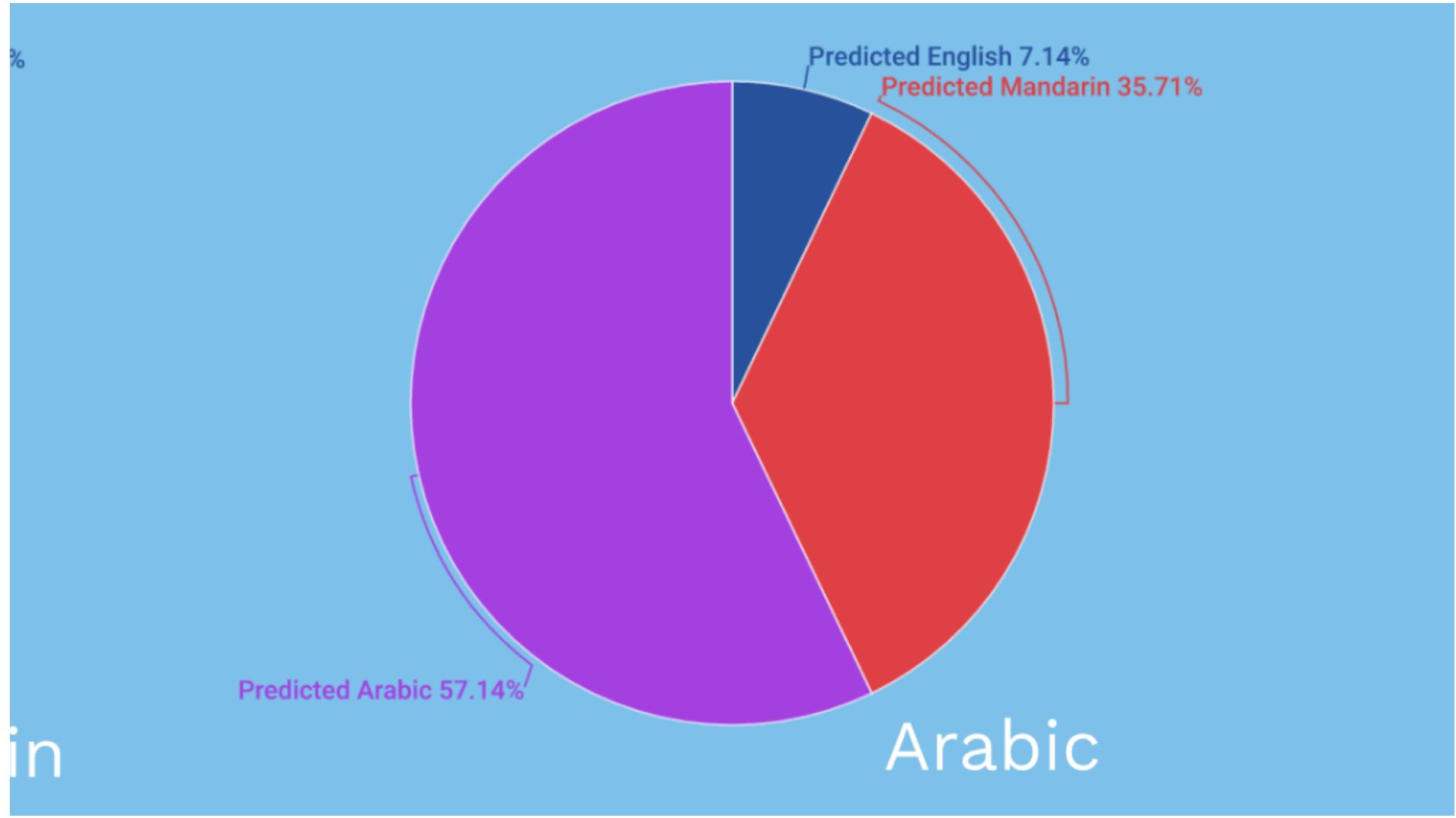


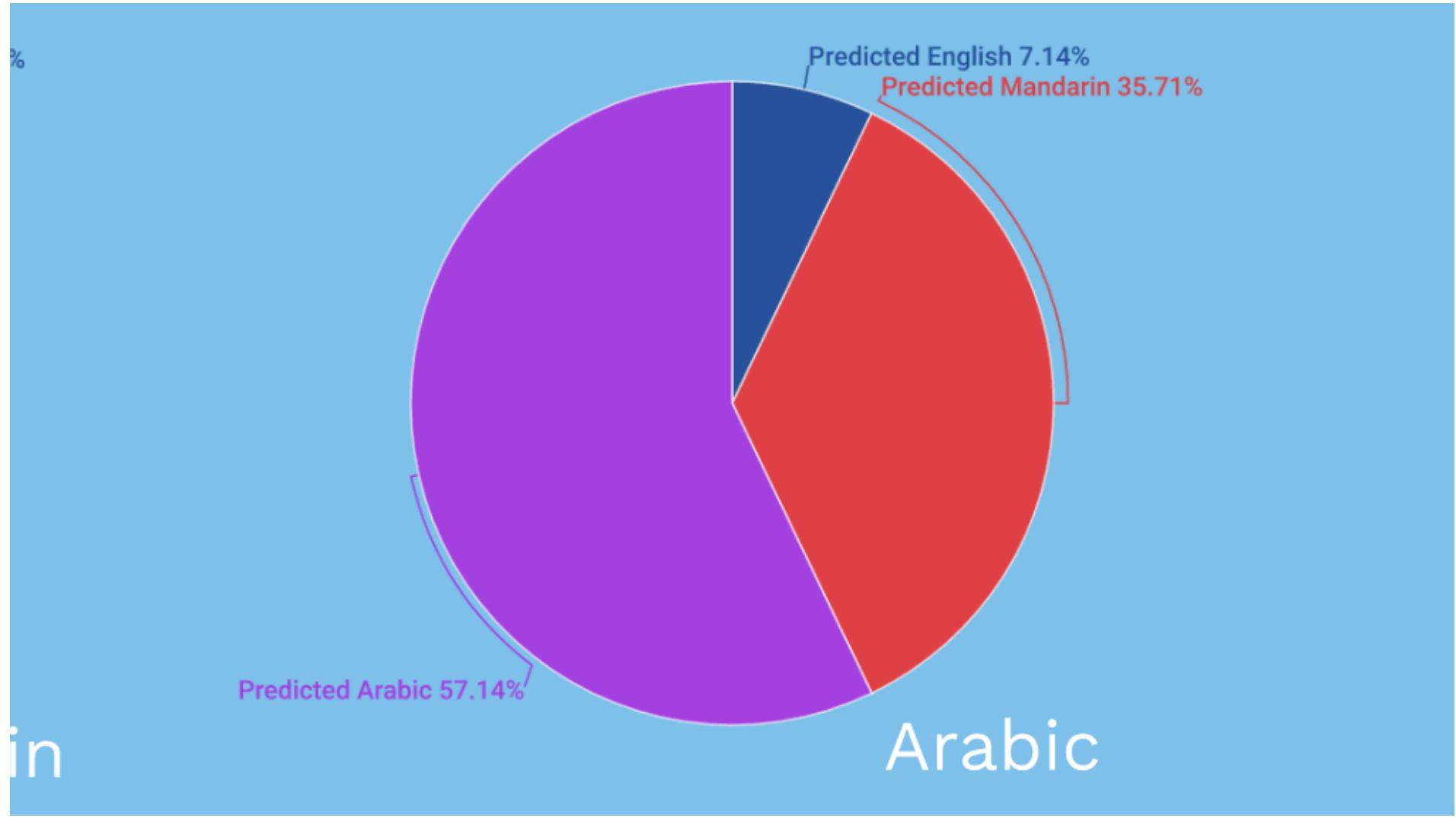
# Results











**77.2%** Accuracy  
over validation set

# DIALECT DETECT

Steven Becerra



# Technologies



K Keras + TensorFlow



BeautifulSoup



# DIALECT DETECT

Steven Becerra





# Future Opportunities

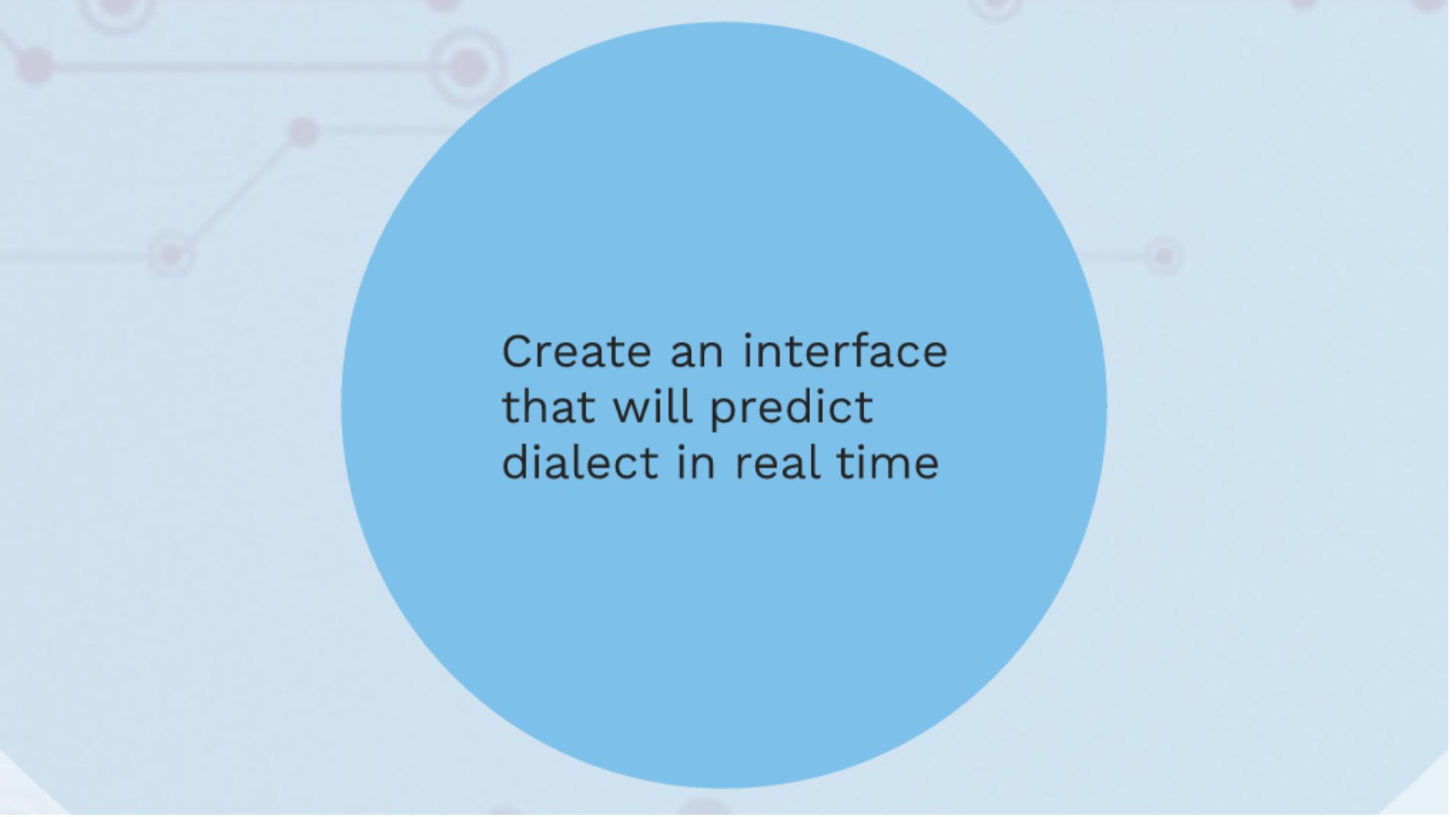
Get More  
Data

Real Time  
Prediction

Thank  
you



Train model on  
audio from outside  
The Speech Accent  
Archive



Create an interface  
that will predict  
dialect in real time

# Thank you

Steven Becerra

<https://github.com/srbecerra>

[srbecerra@gmail.com](mailto:srbecerra@gmail.com)

# DIALECT DETECT

Steven Becerra

