



# Accent Detection in the UK

Mark Streer • METIS (DS/ML) • 2021-10-29

# Dialects, Accents & Algorithmic Bias

SCIENTIFIC  
AMERICAN.

## Speech Recognition Tech Is Yet Another Example of Bias

Siri, Alexa and other programs sometimes have trouble with the accents and speech patterns of people from many underrepresented groups

**Study finds that even the best speech recognition systems exhibit bias**

VentureBeat

The New York Times

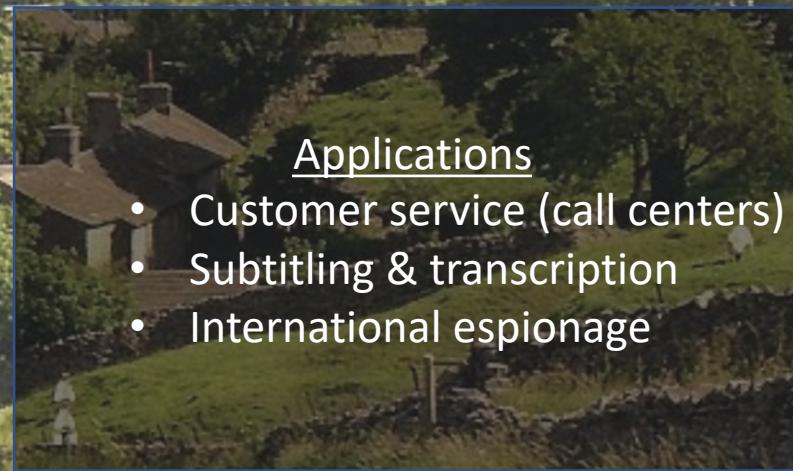
## *There Is a Racial Divide in Speech-Recognition Systems, Researchers Say*

Technology from Amazon, Apple, Google, IBM and Microsoft misidentified 35 percent of words from people who were black. White people fared much better.

# A Different Solution

Specific language models for specific accents/dialects

1. Pretrain models on data from single accent (or group)
2. Determine speaker's accent based on first utterances (0~3 s)
3. Process speech using selected model



# The Data

- Open-source Multi-speaker Corpora of English Accents in the British Isles
  - Created by Google Research ([Demirsahin et al. 2020](#))
  - Raw data available at [OpenSLR.org](#)
- 1 entry = 3~10s spoken sentence by single speaker + transcript
- >18k entries



# The Data

- Open-source Multi-speaker Corpora of English Accents in the British Isles
  - Created by Google Research ([Demirsahin et al. 2020](#))
  - Raw data available at [OpenSLR.org](#)
- 1 entry = 3~10s spoken sentence by single speaker + transcript
- >18k entries
- Six groups of dialects/accents
  1. Southern (English)
  2. Northern (English)
  3. Scottish
  4. Welsh
  5. Midlands (English)
  6. Irish



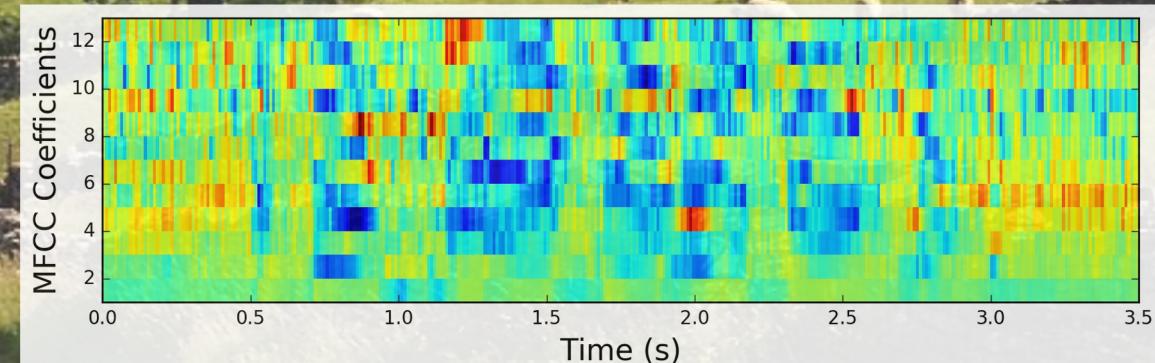
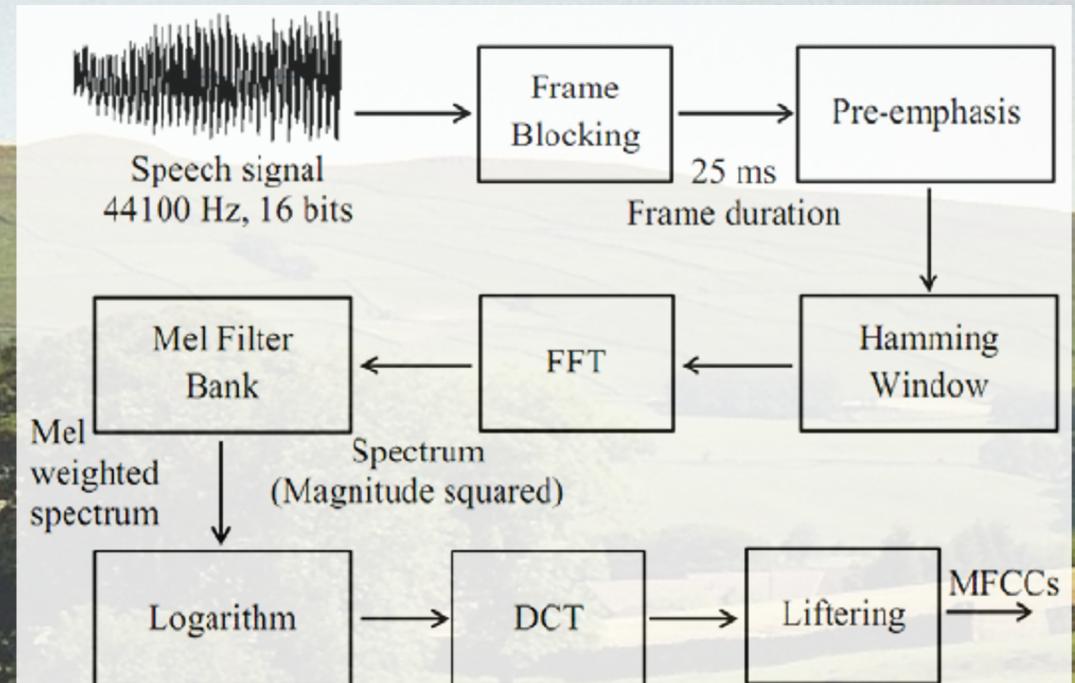
# The Data

- Open-source Multi-speaker Corpora of English Accents in the British Isles
  - Created by Google Research ([Demirsahin et al. 2020](#))
  - Raw data available at [OpenSLR.org](#)
- 1 entry = 3~10s spoken sentence by single speaker + transcript
- >18k entries
- Six groups of dialects/accents
  1. Southern (English)
  2. Northern (English)
  3. Scottish
  4. Welsh
  5. Midlands (English)
  6. Irish



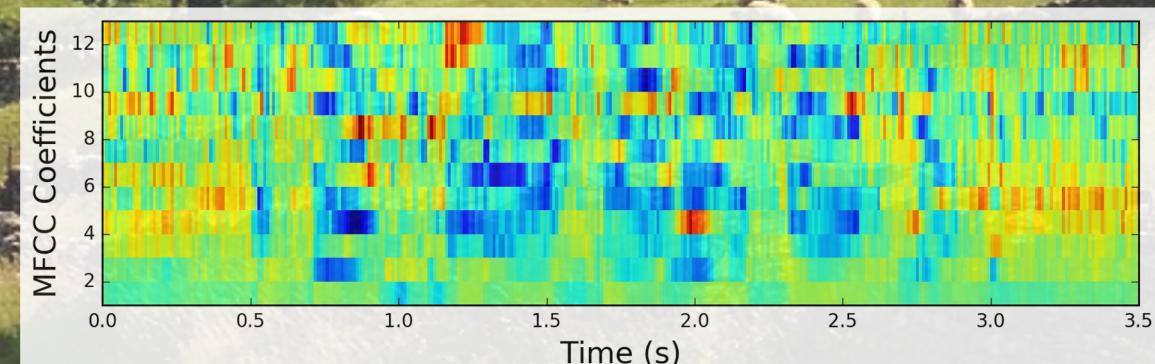
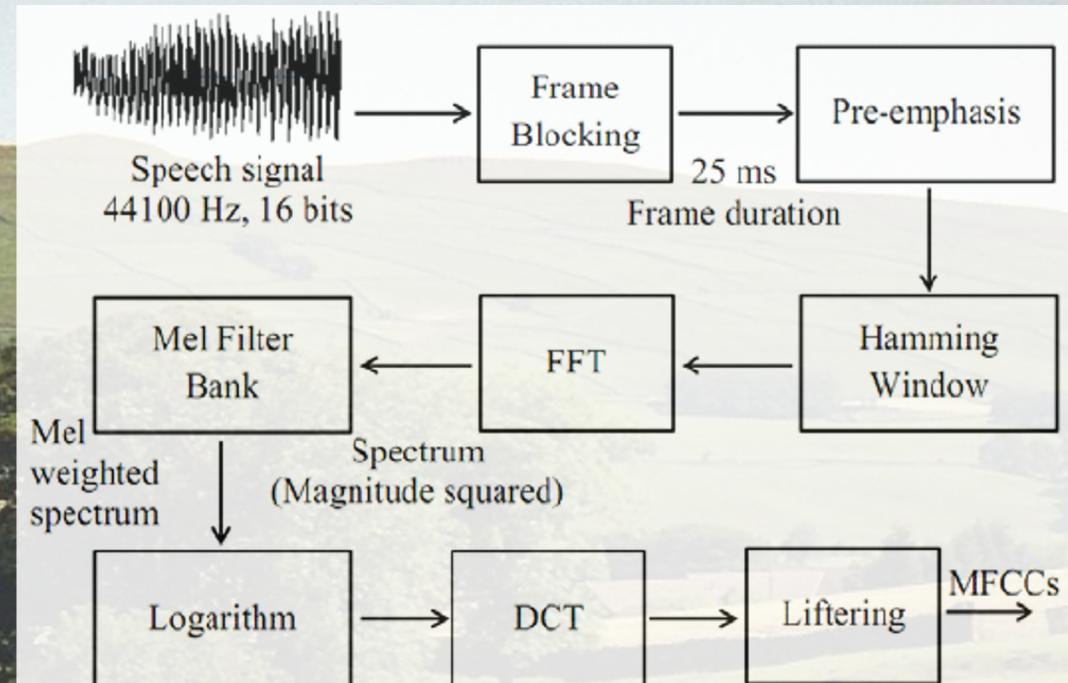
# The Features

- MFCC
  - mel-frequency cepstrum coefficients ('most frequently chosen characteristic')
  - Capture sound energy in frequency bands
  - n=16 typical for speech processing
- MFCC $\Delta$ , MFCC $\Delta^2$ 
  - 1<sup>st</sup>, 2<sup>nd</sup> derivatives of MFCC
  - Capture changes in MFCCs
  - Same n as MFCC



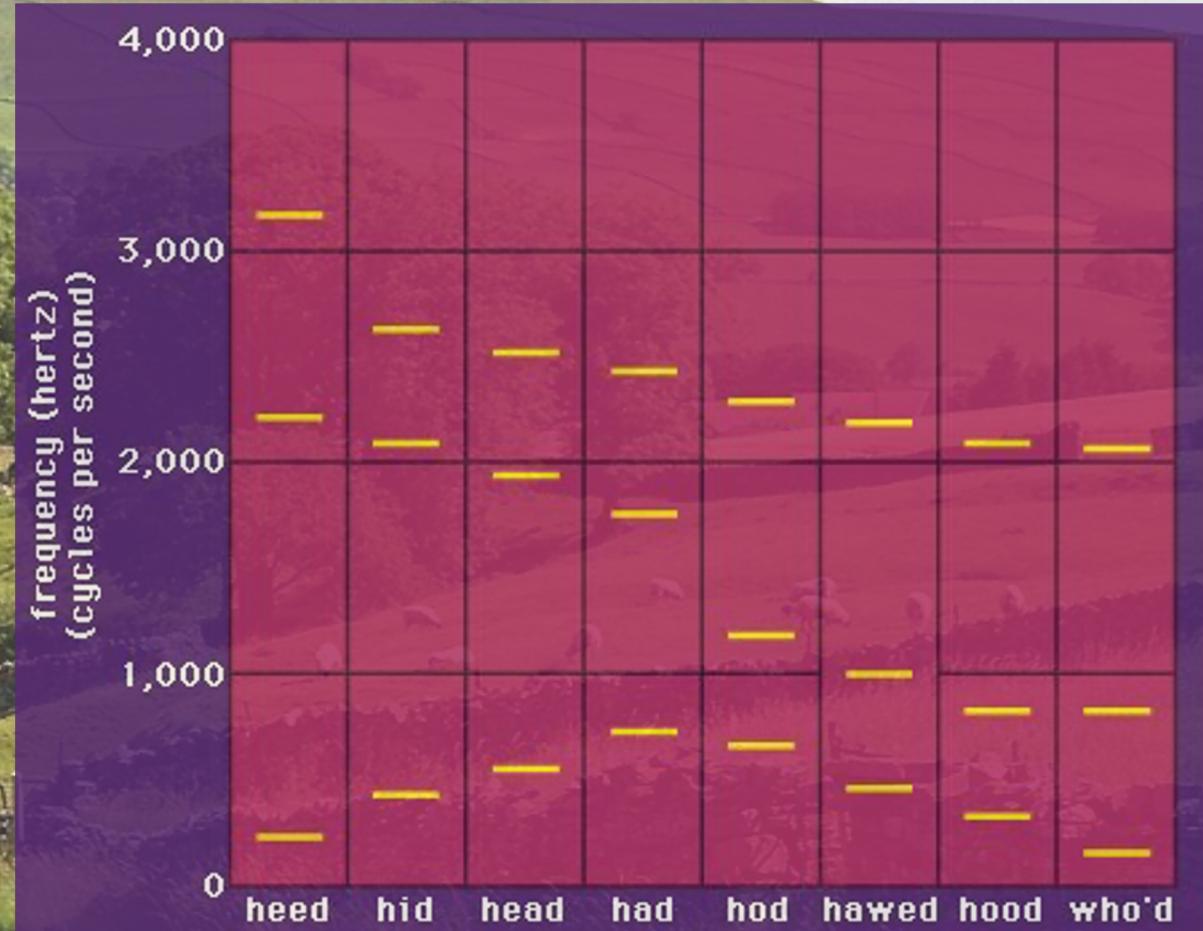
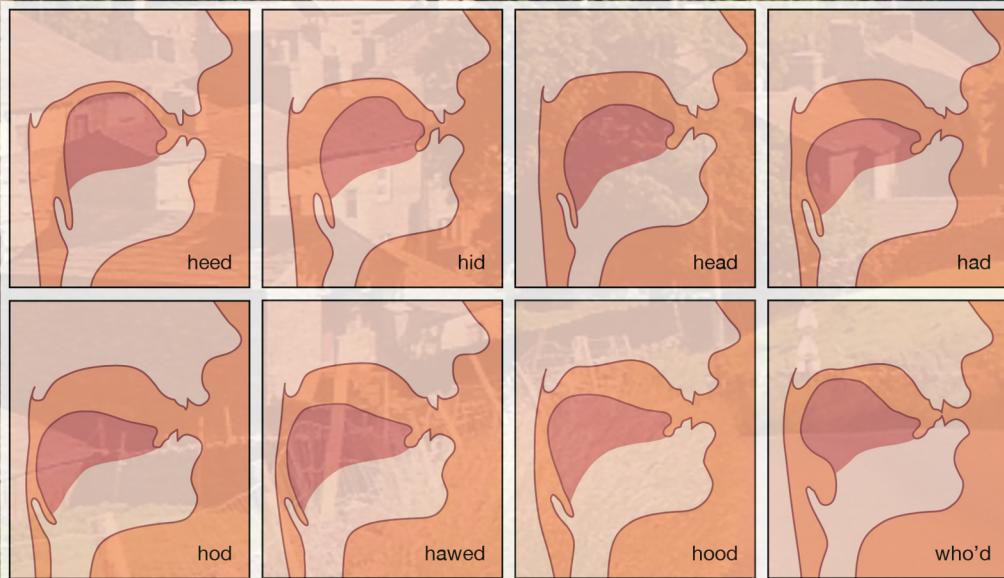
# The Features

- MFCC
  - mel-frequency cepstrum coefficients ('most frequently chosen characteristic')
  - Capture sound energy in frequency bands
  - n=16 typical for speech processing
- MFCC $\Delta$ , MFCC $\Delta^2$ 
  - 1<sup>st</sup>, 2<sup>nd</sup> derivatives of MFCC
  - Capture changes in MFCCs
  - Same n as MFCCs **didn't raise scores**



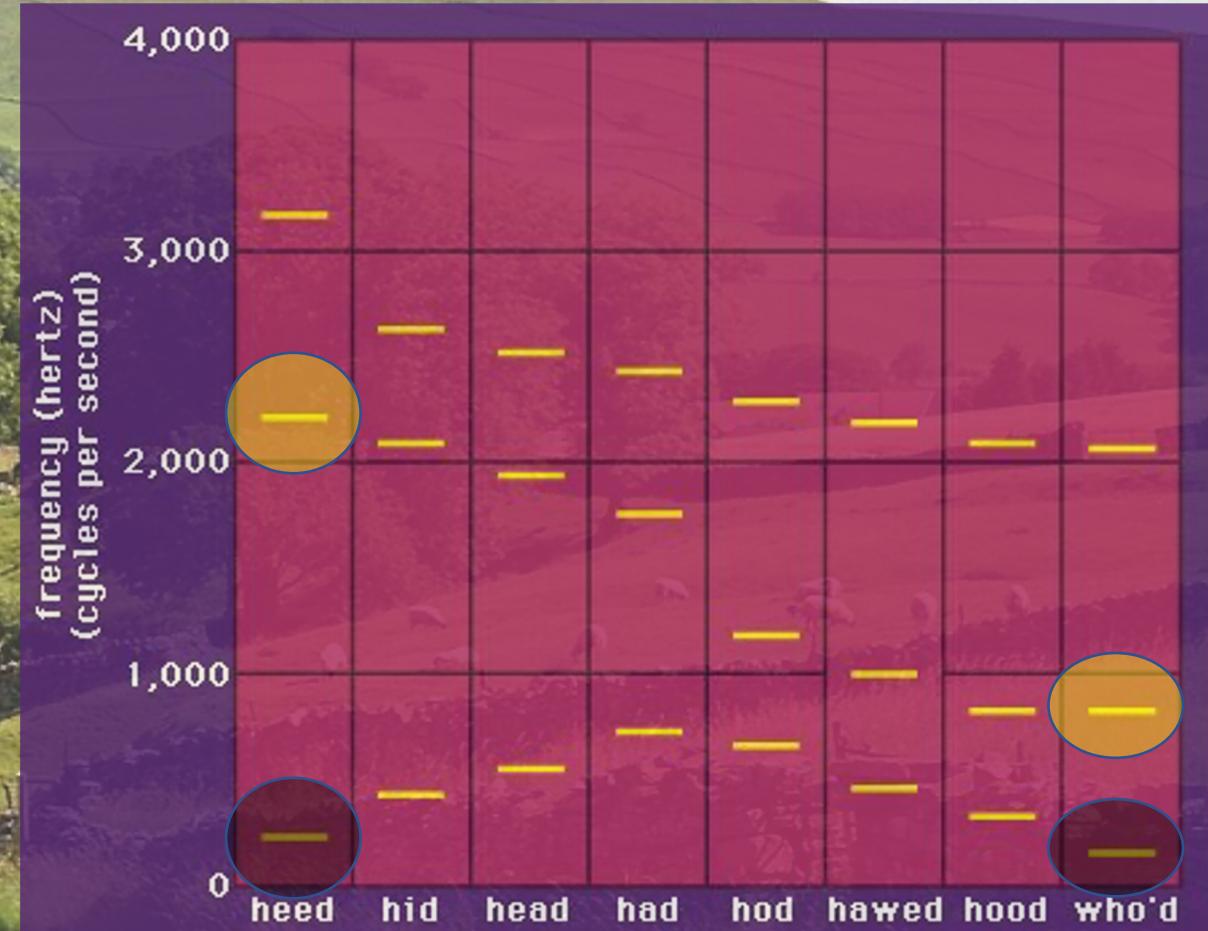
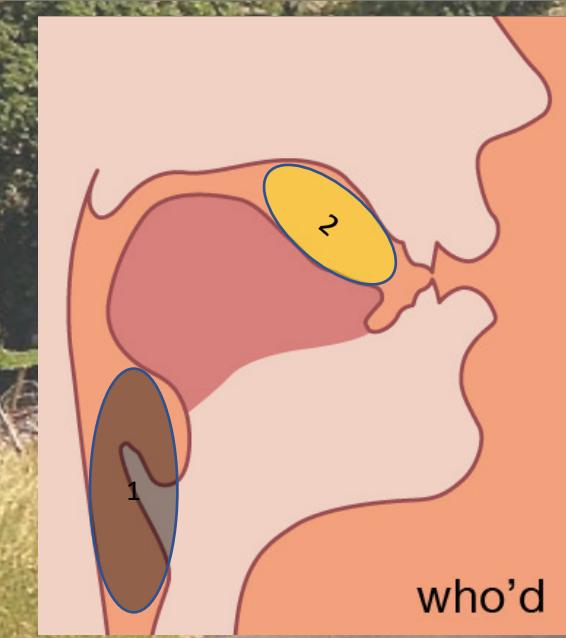
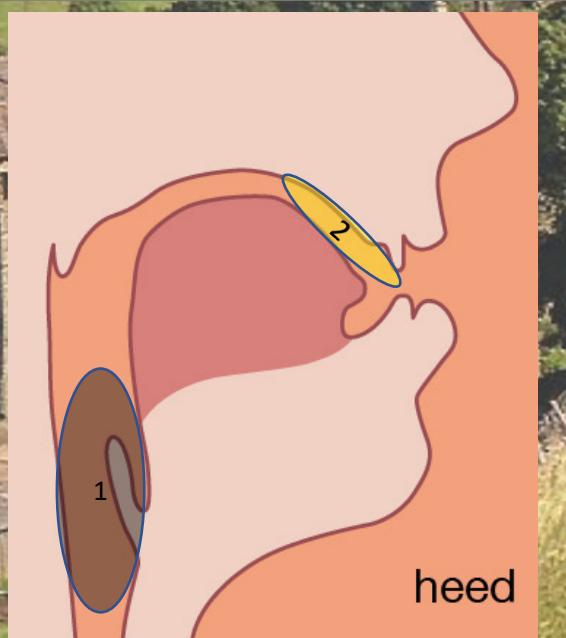
# The Features

- Formants
  - $F_1, F_2, F_3, F_4$
  - Standardize by speaker (divide by  $F_0$ )
  - Mean, median over recording



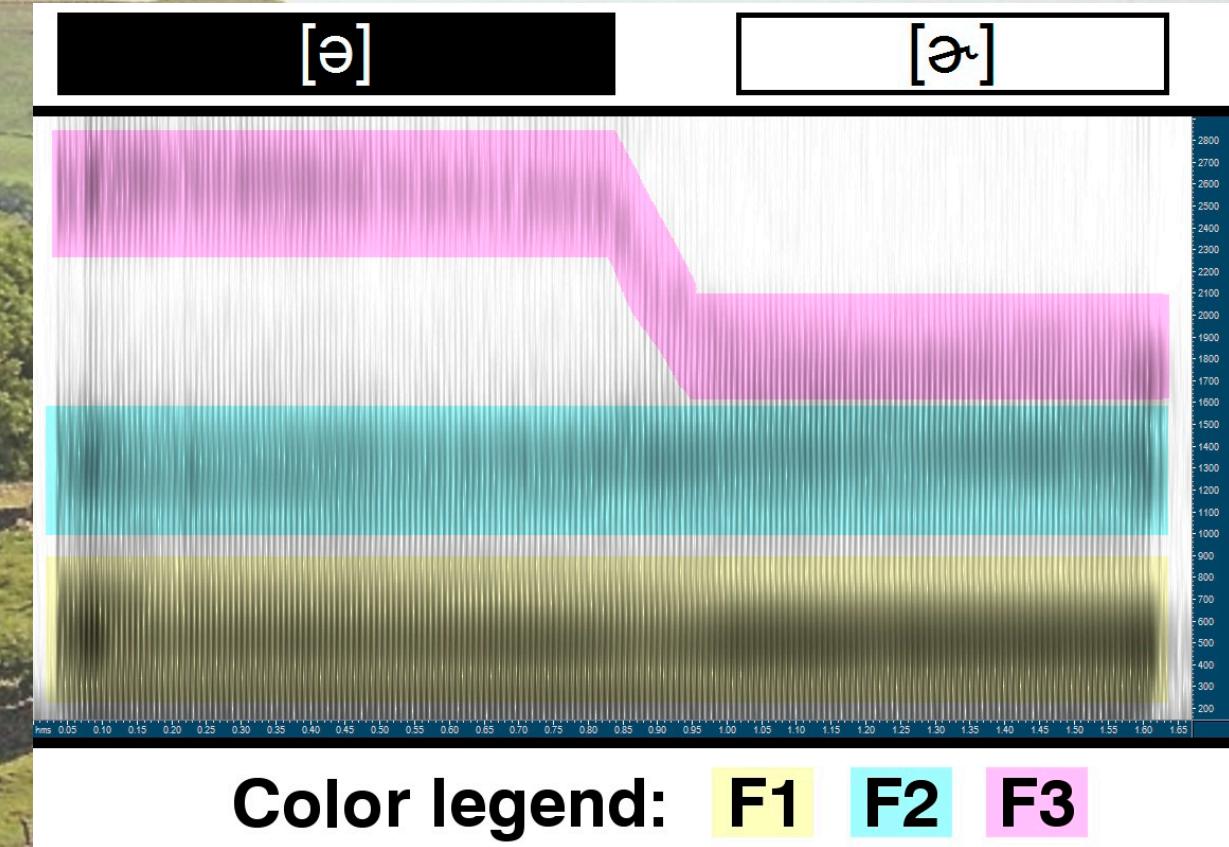
# The Features

- Formant ratios:  $F_2 \div F_1$ 
  - Vowel pronunciation ('raised-ness')
  - High = tend to be close, front
  - Low = tend to be open, back



# The Features

- Formant ratios:  $F_3 \div F_2$   
 $F_3 \div F_1$
- Rhoticity ('r-ness')
  - High = Less rhotacized
  - Low = More rhotacized
- Lip rounding
  - High = Less rounding
  - Low = More rounding
  - ( $F_3$  itself better indicator)



# Model Performance

Multi-class OVR						
Dialect	KNN5 (k=5)	Logistic (C=0.1)	DTC16 (max_depth=16)	RF100 (n_estimators = 100)	XGBoost (params via GridSearchCV)	NB
F <sub>1</sub>	Southern	0.53	0.74	0.53	0.66	0.57
	Northern	0.16	0.35	0.17	0.05	0.75
	Welsh	0.30	0.54	0.34	0.46	0.85
	Scottish	0.21	0.19	0.26	0.18	0.74
	Mean Accuracy (weighted)	0.38	0.55	0.39	0.45	0.37

# Model Performance

Multi-class OVR (class_weight='balanced')						
Dialect	KNN5	Logistic	DTC16	RF100	XGBoost	NB
Southern	0.53	0.74 ↘0.70	0.53 ↗0.56	0.66 ↗0.67	0.57	0.68
Northern	0.16	0.35 ↗0.41	0.17 ↗0.18	0.05 ↗0.28	0.75	0.20
Welsh	0.30	0.54 ↘0.52	0.34 ↘0.17	0.46 ↘0.15	0.85	0.00
Scottish	0.21	0.19 ↗0.25	0.26 ↗0.35	0.18 ↗0.29	0.74	0.00
Mean Accuracy (weighted)	0.38	<u>0.55</u>	0.39	0.45	<u>0.55</u>	0.37

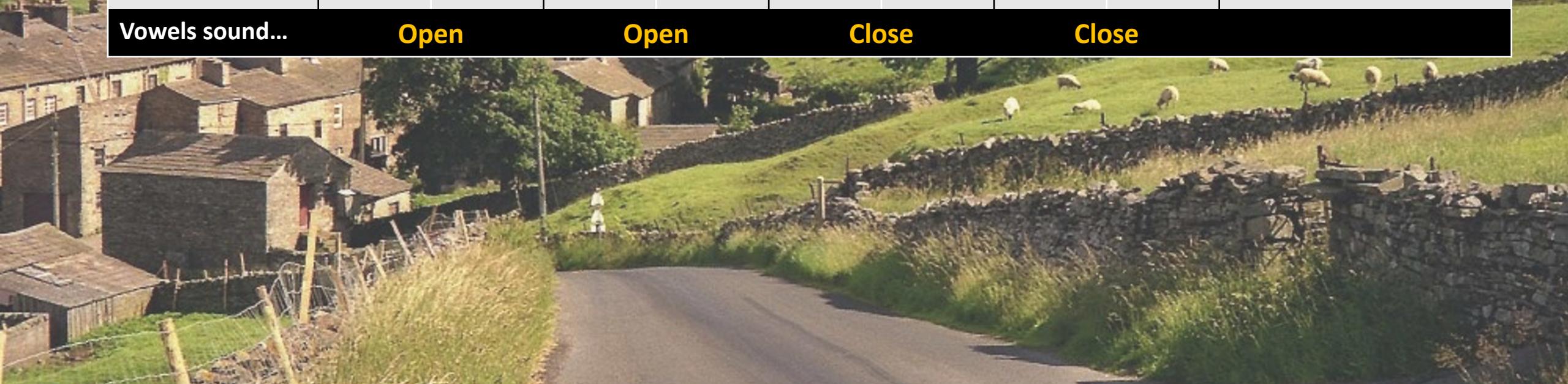
Ranking: **XGBoost** ≈ **Logistic** >> **RF100** >> **DTC16** > **KNN5** > **NB**

# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)

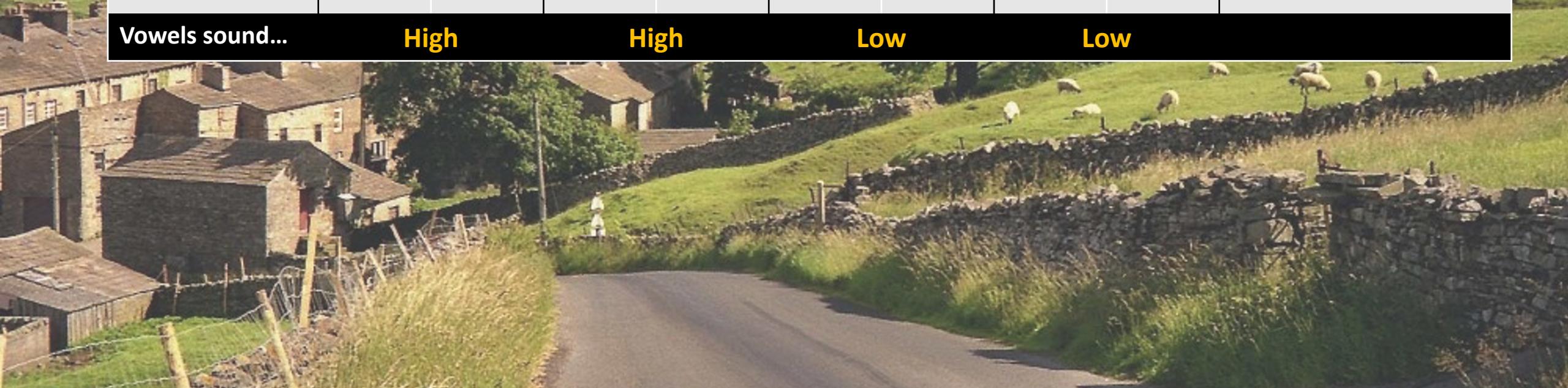
# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)
Vowels sound... Open Open Close Close									



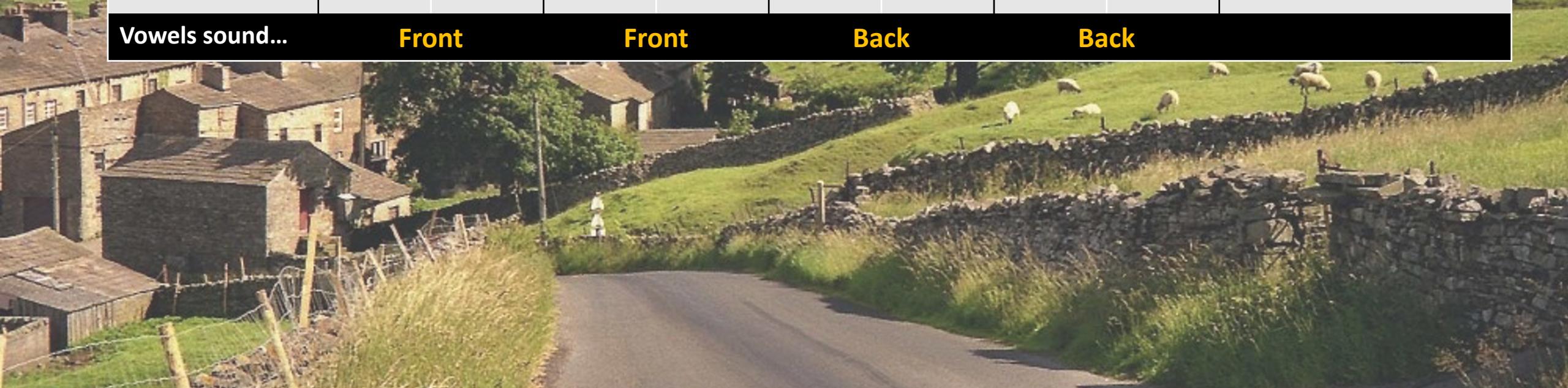
# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)
Vowels sound...	High		High		Low		Low		



# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)
Vowels sound...	Front		Front		Back		Back		

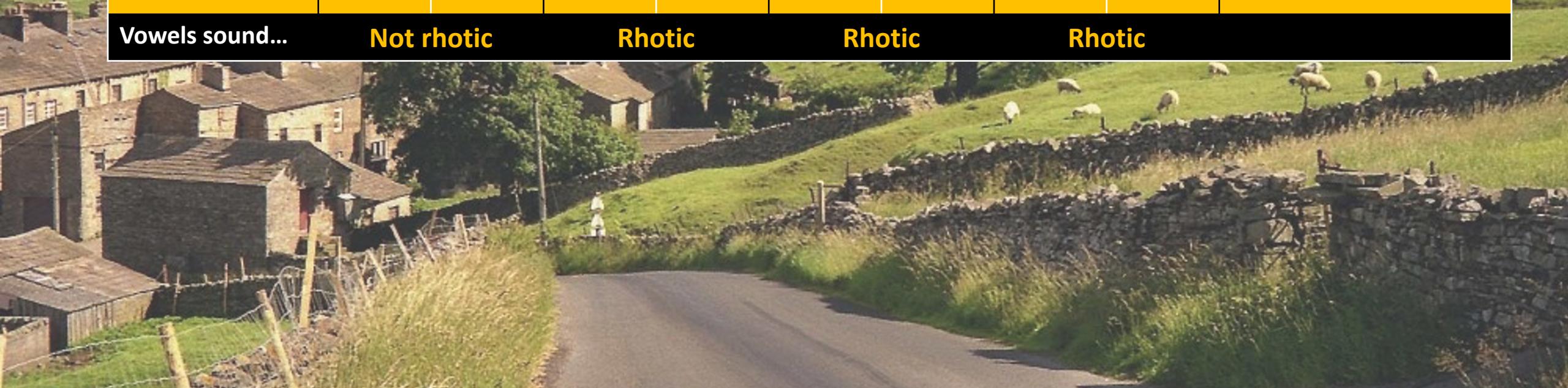


# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)
Vowels sound...	Unrounded		Rounded		Rounded		Unrounded		

# Feature Importance

(Logistic)	Northern		Scottish		Southern		Welsh		Related to
F3/F1_median	-0.82	#4	-2.00	#2	1.25	#1	1.57	#2	Rhoticity (+rounding)
F1_median (F1_mean)	-1.28	#1	-1.68 (-2.33)	#3 (#1)	0.89	#4	0.25 (2.21)	#3 (#1)	Vowel height
F2/F1_mean	+		1.62	#4	-0.95	#2	-0.97	#4	Vowel compactness
F3_mean	0.98	#2	—		-0.93	#3	+		Lip rounding
F3/F2_median	-0.84	#3	+		+		+		Rhoticity (+rounding)
Vowels sound...	Not rhotic		Rhotic		Rhotic		Rhotic		



# Future Directions

- Run modeling separately by gender
- Check only first (or last) 1~3 sec of recordings
- Include more audio features (e.g., jitter, shimmer)
- Ethical ramifications of profiling