Classification of Vowel Sounds by Frequency

Henry N, Gavin J, Kaitlyn S, Kyle Z, Laura K

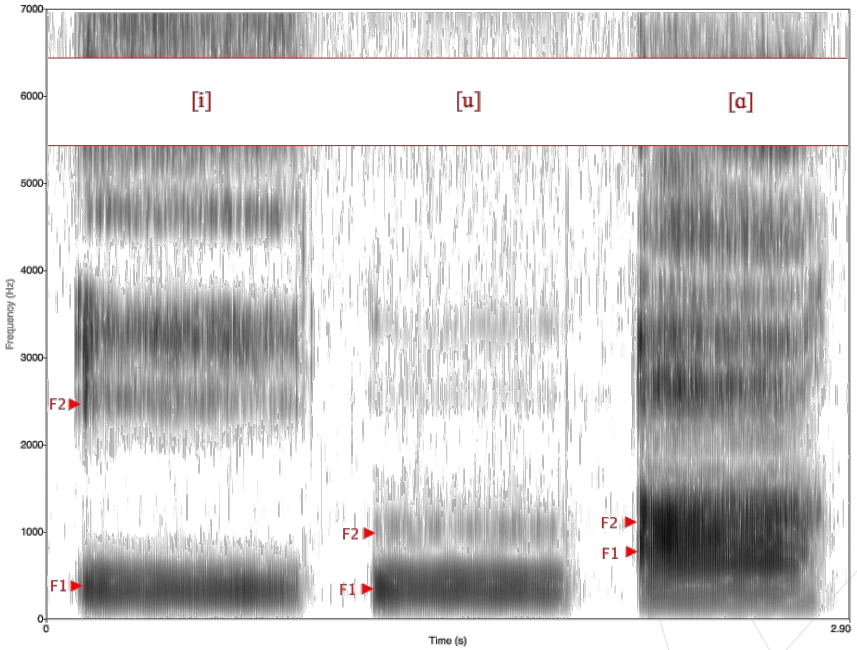
Introduction: Data

- Data from Peterson & Barney (1952) study on acoustic analysis
- Downloaded from phonTools R package
- 1520 vowel sounds from 76 speakers
- 4 sound frequencies extracted from each measurement
- **Sounds labeled by their symbols in the X-SAMPA phonetic library:**

- "u" is the "oo" in "boot"
- "i" is the "ee" in "meet"
- "E" is the "e" in "bet"
- "A" is the "a" in "father"
- "{" is the "a" in "trap"
- "3'" is the "u" in "bud"
- "U" is the "oo" in "foot"
- "O" is the "o" in "off"
- "V" is the "u" in "gut"
- "I" is the "i" in "kit"

Introduction: Background

- Can frequency measurements from recordings determine which vowels are being spoken?
- Which models best predict the gender of the sample's speaker and which vowel they are saying?

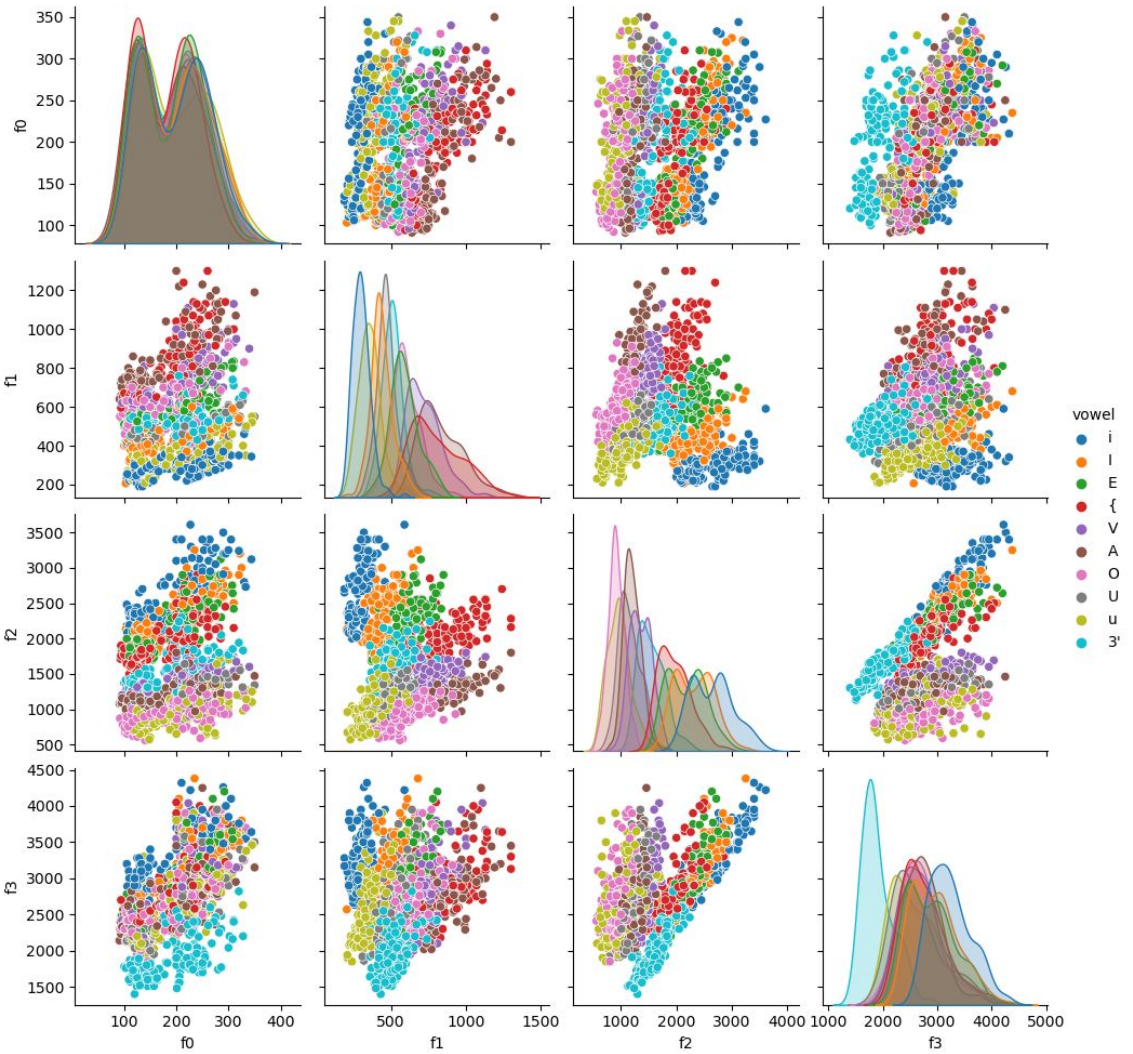


Features

- **Sex** (male or female)
- **F0: fundamental frequency** of a sound wave. The fundamental frequency is the lowest frequency of a periodic waveform, and it corresponds to the perceived pitch of a sound.
- **F1 (First Formant):** This is the lowest frequency formant and is primarily influenced by the tongue height. In general, as the tongue moves higher in the oral cavity (closer to the roof of the mouth), f1 increases.
- **F2 (Second Formant):** This formant is influenced by tongue advancement or frontness/backness. As the tongue moves forward in the oral cavity (towards the front), f2 increases.
- **F3 (Third Formant):** This formant is influenced by additional tongue movements and contributes to the overall vowel quality. The exact factors influencing f3 can be more complex and language-specific.

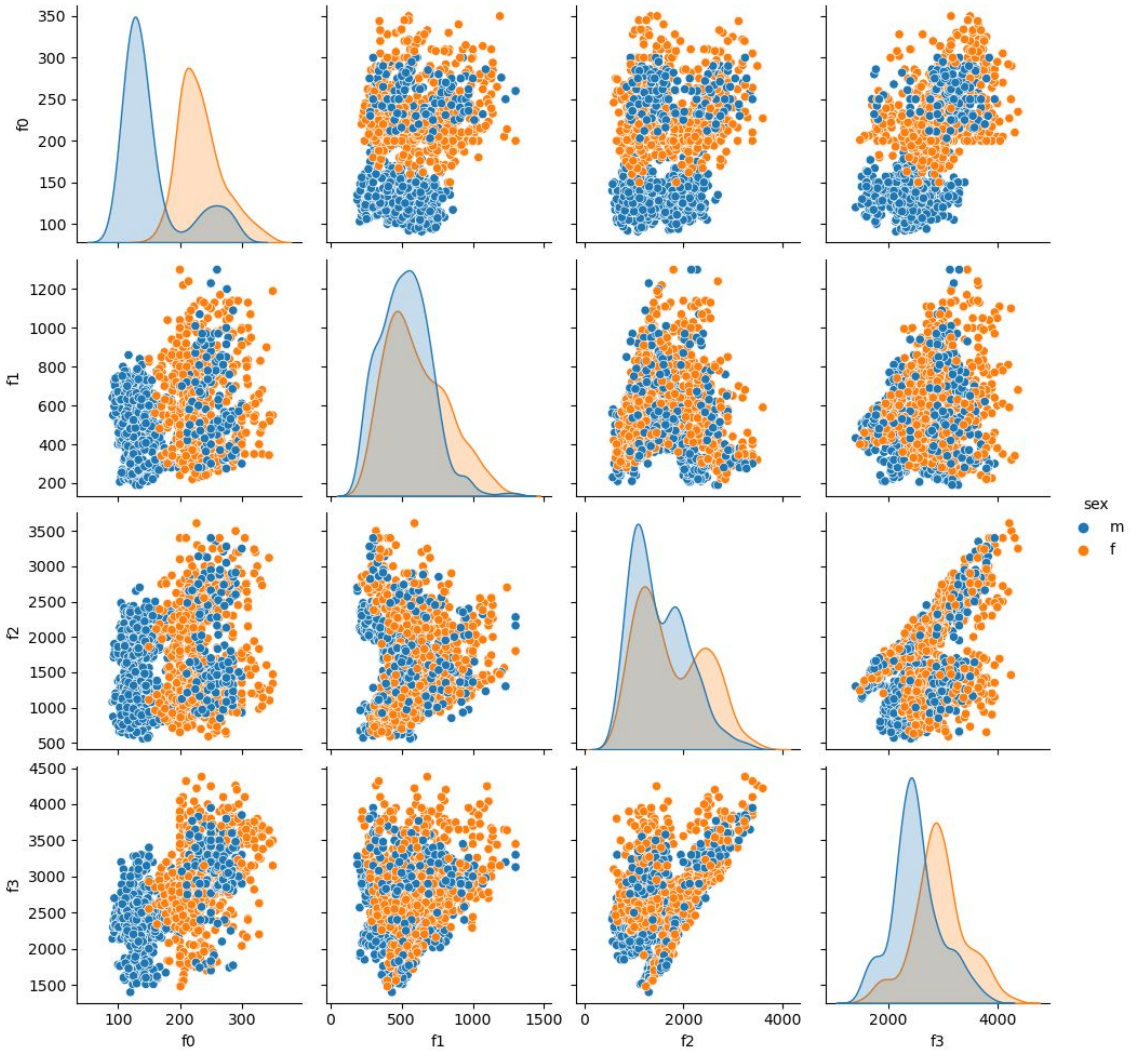
Vowel Pairplot

- High f0 collinearity indicates that this is not a good predictor
- Other features show clustering
- Lots of overlap - models need many features for high accuracy



Gender Pairplot

- f0 feature demonstrates significantly less overlap in the data - indicates that it is a good feature for predicting gender
- Other features display more overlap



Preprocessing

- Changed sex feature to numeric 0/1
- Converted from R dataframe to .csv to Pandas dataframe
- Normalized values using standardization in Scikit-learn
- For sex predictions added one-hot encoded features for each vowel

	type	sex	speaker	vowel	repetition	f0	f1	f2	f3
0	m	0	1	i	1	160	240	2280	2850
1	m	0	1	i	2	186	280	2400	2790
2	m	0	1	I	1	203	390	2030	2640
3	m	0	1	I	2	192	310	1980	2550
4	m	0	1	E	1	161	490	1870	2420
...	...	...	...	...	...	...	...	...	...
1515	c	1	76	U	2	322	610	1550	3400
1516	c	1	76	u	1	345	520	1250	3460
1517	c	1	76	u	2	334	500	1140	3380
1518	c	1	76	3'	1	308	740	1850	2160
1519	c	1	76	3'	2	328	660	1830	2200

1520 rows × 9 columns

Modeling Results - SVM

- Strongest model:
  - For classifying gender: SVC(kernel = “linear”, C = 100) ; **0.90789**
  - For classifying vowel: SVC(kernel = “linear”, C = 1) ; **0.88158**

Modeling Results - Logistic Regression

- Strongest model:
  - For classifying gender: LogisticRegression(max_iter=1000, C=1);

    0.86842
  - For classifying vowel: LogisticRegression(max_iter=1000, C=1000);

    0.88158

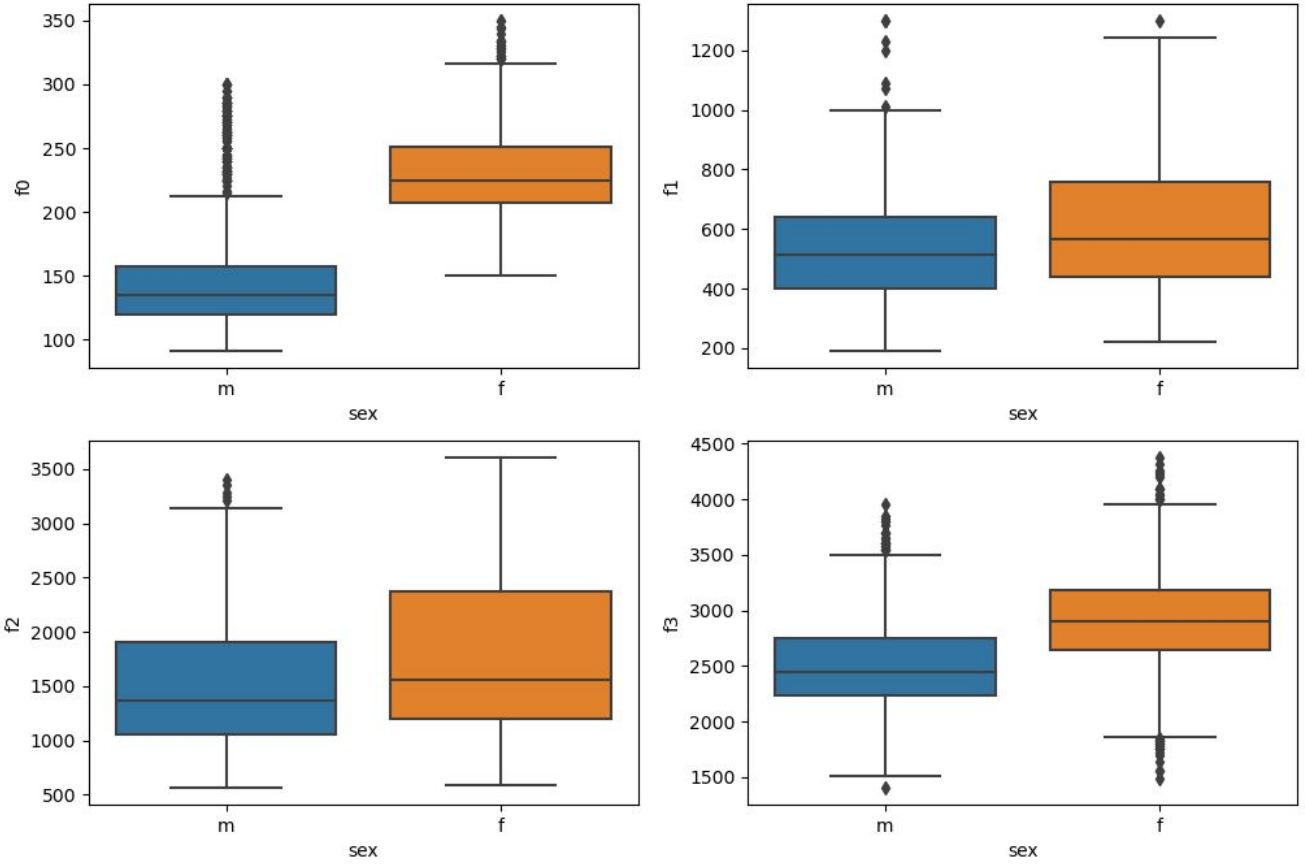
Modeling Results - Decision Tree

- Strongest model:
  - For classifying gender: DecisionTreeClassifier(criterion='entropy', max_depth=1); **0.88158**
  - For classifying vowel: DecisionTreeClassifier(criterion='entropy', max_depth=7); **0.80921**

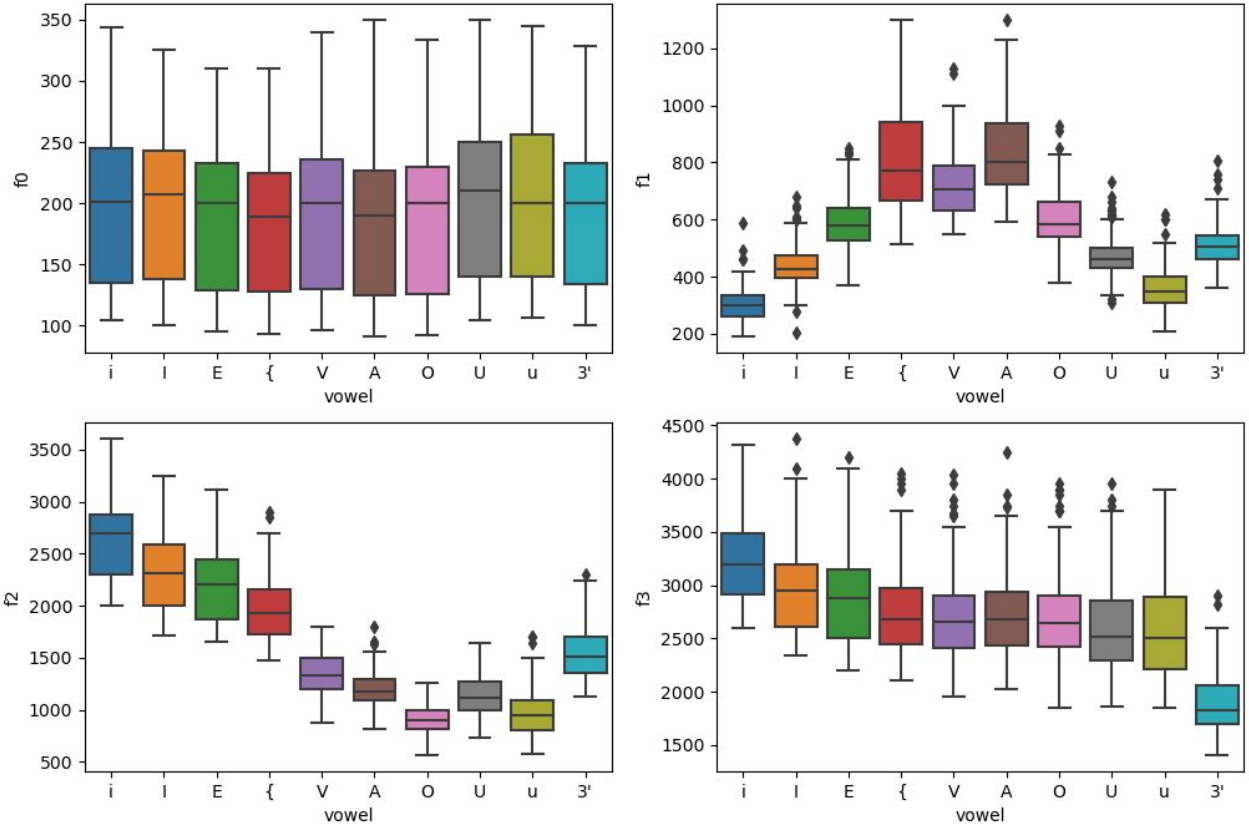
Modeling Results - KNN

- Strongest model:
  - For classifying gender: DecisionTreeClassifier(criterion='entropy', max_depth=1); **0.88158**
  - For classifying vowel: DecisionTreeClassifier(criterion='entropy', max_depth=7); **0.82237**

Variable Importance - Gender



Variable Importance - Vowels



Conclusions

Best models according to GridSearchCV:

- Classifying by gender/sex: KNeighborsClassifier(metric='euclidean', neighbors=5) ; **0.91447**
- Classifying by vowel: SVC(C=1, kernel='linear') ; **0.88158**

Best features:

- Classifying by gender/sex: **f0/pitch**
- Classifying by vowel: **f1/f2 formants**

Applications

- Record our group members saying each of the vowels we tested on and measure each frequency
- Test each of the models on the true data and see which ones can accurately predict gender and the vowels correctly
- Technology is useful for speech recognition, dialect classification, transcription, etc.

Thanks!