Vowel Recognition

Linear Systems Lab: ECE 3151-50

Eric Adamany, Robert Campbell, Christian Litalien

12/5/2019

Introduction:

The purpose of this lab was to learn how to create a classification tree in order to implement a vowel recognition system. This was done by observing the energy graphs of the 75 total vowels and determining the separations between the different vowel groups. Once the separations were found for each vowel, the tree was constructed and tested using a function given.

Results:

A. When creating our classification tree, we stepped through each graph to find the best separation between the vowels. First we looked for a separation between all the vowels and found that [E(7) E(9)] separate AIO from EU. Next stepped through again looking at only E and U. It was found that [E(4) E(12)] gave a good separation between the two vowels. Next we looked for a separation between AIO and found that [E(3) E(10)] divided I from AO. Finally we found the separation for AO to be [E(1) E(13)]. These values were used to create our classification tree and classify\_vowel.m. The full tree can be seen attached to the end of the report.

B. With our original decision tree, nearly all of the training data was classified appropriately except for student #2’s E and U’s. We used that information from the confusion matrix to separate out student #2 using [E(1) E(2)] and then used [E(1) E(11)] to separate E and U for student #2 while still using [E(4) E(12)] to separate those vowels for students #1 and #3. After these changes, the training data only miscategorized 3 vowels (see Training Data matrix below).

In general, our classifier works pretty well with some difficulty in classifying U and I. This difficulty is much more pronounced in the test data than it is in the training data (see Test Data matrix below). This could be due to any number of things, including uncontrolled changes in the recording environments for the two samples, as the training samples were recorded days apart from the test samples.

Training Data

group student #1 student #2 student #3

a e i o u a e i o u a e i o u a e i o u

---|------------------|------------------|------------------|------------------|

a | 74 0 1 0 0 | 25 0 0 0 0 | 24 0 1 0 0 | 25 0 0 0 0 |

e | 0 75 0 0 0 | 0 25 0 0 0 | 0 25 0 0 0 | 0 25 0 0 0 |

i | 2 0 73 0 0 | 0 0 25 0 0 | 2 0 23 0 0 | 0 0 25 0 0 |

o | 0 0 0 75 0 | 0 0 0 25 0 | 0 0 0 25 0 | 0 0 0 25 0 |

u | 0 1 0 0 74 | 0 1 0 0 24 | 0 0 0 0 25 | 0 0 0 0 25 |

Test Data

group student #1 student #2 student #3

a e i o u a e i o u a e i o u a e i o u

---|------------------|------------------|------------------|------------------|

a | 41 0 0 2 2 | 12 0 0 2 1 | 15 0 0 0 0 | 14 0 0 0 1 |

e | 0 45 0 0 0 | 0 15 0 0 0 | 0 15 0 0 0 | 0 15 0 0 0 |

i | 1 0 24 7 13 | 0 0 15 0 0 | 1 0 1 7 6 | 0 0 8 0 7 |

o | 0 0 2 43 0 | 0 0 0 15 0 | 0 0 0 15 0 | 0 0 2 13 0 |

u | 0 15 0 0 30 | 0 12 0 0 3 | 0 3 0 0 12 | 0 0 0 0 15 |

Market Analysis:

C. Our group is composed of males from Arizona, Missouri, and Texas so our primary target market would typically be males from the Midwest and Southwest United States. In the US census of 2018, the Midwest had a population of 68,308,774 persons while the Southwest had a population of 124,753,948. To determine the total amount of people in our market size, we can sum the populations and obtain 193,062,722 persons.

Due to its limited capabilities and specific target market, the cost of our product would be $20. Out of the 193,062,722 persons, we can presume that the majority of people could purchase the product. According to a [2015 survey](https://www.thesimpledollar.com/dont-eat-out-as-often-188365/), Americans spend an average of $20 per week eating out for lunch and in a [2018 survey](https://www.foxnews.com/food-drink/millennials-are-spending-more-money-on-coffee-than-retirement-plans), the average American will spend $1,100 a year or $3 a day on coffee. With an average total of $41 per week on eating out and coffee, it could be presumed that a majority of our target audience could afford to purchase our product.

While comparing our speech recognition system to our market competition, [Dragon Speech Recognition](https://www.capterra.com/p/21494/Dragon-NaturallySpeaking/#about) and [Braina](https://www.capterra.com/p/145652/Braina/#pricing) will be some of the top competitors. Dragon Speech Recognition pricing starts at $199.99 as a one time payment, per user. It has speech recognition features such as: audio capturing, automatic transcription, customizable, macros, and speech-to-text analysis. The speech recognition system, Braina, starts at $39.99 per year, per user. It has speech recognition features such as: audio capturing, automatic transcription, concatenated speech, customizable macros, and is multi-language. Since our product lacks most of these premium features offered or a customer/technical support channel, we cannot compete at the top of the market with big name brands. We offer a low-complexity voice recognition system that at a low price that is obtainable to a majority of our target market without exceeding the price of coffee per week.

Conclusion:

In conclusion, the purpose of this lab was to provide the students with the opportunity to develop a speaker-independent vowel recognition system which will distinguish between 5 long vowel sounds. The students will also come to understand the importance of the Fourier Spectrum for developing speech recognition systems, the use of a metric space for performing classification, how to gain practical knowledge on the relationship between commercial viability and system performance. In the Vowel Recognition lab, was asked to record 25 vowel sounds for each of the long vowels: ‘a’,’e’,’i’,’o’,’u’. By combining the vowel sounds we calculated their respective spectral energies and plotted them over an appropriate frequency range and found the best plots to use for training data. The vowel separation and decision points can be seen directly from the Multi-Level Classifier Tree. After generating our first confusion matrix, we noticed the 'u's from student #2 were not being classified correctly. To resolve the issue, we added an additional decision point to separate the one speaker from the rest and categorize that student's 'e's and 'u's independently from the others. This resulted in nearly all of our training vowels being classified correctly. Following the classification of the vowel sounds, we performed a Market Analysis of our product. We could deduce that our primary target market would be males or individuals with lower voices in the Midwest and Southwestern regions of the United States.

Bibliography:

<https://www.thesimpledollar.com/dont-eat-out-as-often-188365/>

<https://www.foxnews.com/food-drink/millennials-are-spending-more-money-on-coffee-than-retirement-plans>

<https://www.capterra.com/p/21494/Dragon-NaturallySpeaking/#about>

<https://www.capterra.com/p/145652/Braina/#pricing>

