Q1 Run your code 10 or more times with random initializations. What range do you observe for each of precision, recall, and f-score? What is the f-score of the best clustering you found?
Precision range: 0.27627667147206786 – 0.34742745784038426
Recall range: 0.379743488033849 – 0.4362025651196615
F-score for best clustering: 0.3833968334603336

Q2 Run your program on the child, female, and male data separately. You can do this by either having your program temporarily ignore certain lines of input, or by separating out the vowel data into three input files. What are the best precision, recall, and f-score values you get for each of these three inputs? Why are the results different from running the program on all the data at once?

Male:
Precision: 0.5173141367754748
Recall: 0.6102786861242618
F-score: 0.5599641737572771

Female:
Precision: 0.5439570068345113
Recall: 0.6360655737704918
F-score: 0.5864164488806323

Child:
Precision: 0.49136857092009256
Recall: 0.6476659629368989
F-score: 0.5587937664440396

The results for running the program on child, female, and male data separately are better than running the program on all the data at once because there is less variation in terms of F1 and F2 within these demographics. As mentioned above, male speakers have lower pitch than female speakers, who have lower pitch overall than children. Therefore, the F1 and F2 values for these three demographics are closer in range and the vowel clusters will have less overlap if graphed them separately as opposed to all together.

Q3 Modify your code so that F1 and F2 are represented in 'mels' instead of hertz. See the following for a definition of mels: http://en.wikipedia.org/wiki/Mel_scale What effect does this have? Why do you think this happens and what is the cognitive significance of this (hint: think about the shape of the clusters in both spaces)?
This increases the precision, recall, and f-score slightly: precision is ~0.38, recall ~0.43, and F-score ~0.41. I think this is because mels is a "perceptual scale of pitches judged by listeners to be equal in distance from one another," according to the Wikipedia. Looking at the formula to convert hertz to mels, we can see that it uses log– this captures how humans perceive lower frequencies to be more different from each other than higher frequencies with the same

difference in frequencies. Therefore, using mels will expand lower frequencies and compress higher ones, which is better for capturing the clusters of vowels based on how we actually categorize them.

Q4 Here's your chance to be creative! Extend your program in some way to take advantage of the information contained in the first, second, or fourth columns of the data. Explain clearly what changes you made, why, and what effect they had on the accuracy of your program. Submit the new version of your program with a new name.