Speech Phoneme Analysis and Classification

ICS2203 - Statistical NLP Assignment 2025

Miguel Baldacchino
University of Malta

May 2025

Contents

Overview

This assignment's primary focus is the development of a speech classifier for vowel-based phonemes, using formant frequency analysis and k-NN classification.

The corpus used is a subset of the ABI-1 corpus, consisting of: 5 selected british regional accents, with speech samples from 5 male and 5 female speakers per accent, with 3 vowel phonemes per speaker (UH, IH, AE).

Formant frequencies were manually extracted using Praat, by isolating pure vowel segments in each word, and analyzing via spectrogram. Formant tracks (F1, F2, F3) were extracted via the formant analysis tool.

FACULTY OF INFORMATION AND COMMUNICATION TECHNOLOGY

Declaration

Plagiarism is defined as "the unacknowledged use, as one's own work, of work of another person, whether or not such work has been published" (Regulations Governing Conduct at Examinations, 1997, Regulation 1 (viii), University of Malta).

I / We*, the undersigned, declare that the [assignment / Assigned Practical Task report / Final Year Project report] submitted is my / our* work, except where acknowledged and referenced.

I / We* understand that the penalties for making a false declaration may include, but are not limited to, loss of marks; cancellation of examination results; enforced suspension of studies; or expulsion from the degree programme.

Work submitted without this signed declaration will not be corrected, and will be given zero marks.

* Delete as appropriate.

(N.B. If the assignment is meant to be submitted anonymously, please sign this form and submit it to the Departmental Officer separately from the assignment).

Miguel Baldacchino
Student Name


Signature

Student Name

Signature

Student Name

Signature

Student Name

Signature

ICS2203
Course Code

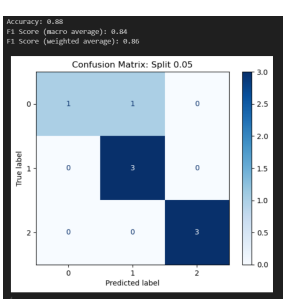
Speech Phoneme Classifier
Title of work submitted

22/05/2025
Date

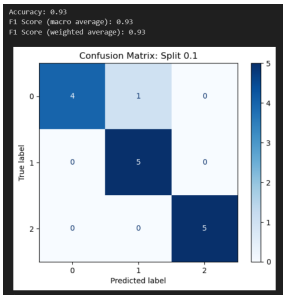
1 Experiments and Results

1.1 Varying Split

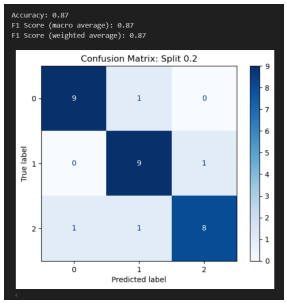
Classifier reached peak performance at 10% test split, with an accuracy and F1 of 0.93, indicating highly accurate predictions and minimal confusion across classes. As the test split increased, performance very slightly declined, with the 25% split having the lowest scores and more dispersion in the confusion matrix, especially between similar phonemes. The 20% split was chosen for the rest of evaluation as it offers performance very close to the best performing split (10%) but allowing a larger and more reliable test set size for analysis.



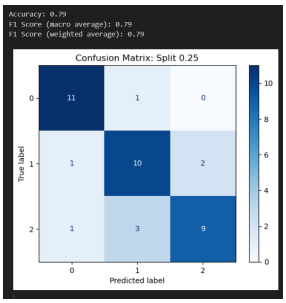
(a) Split 5%



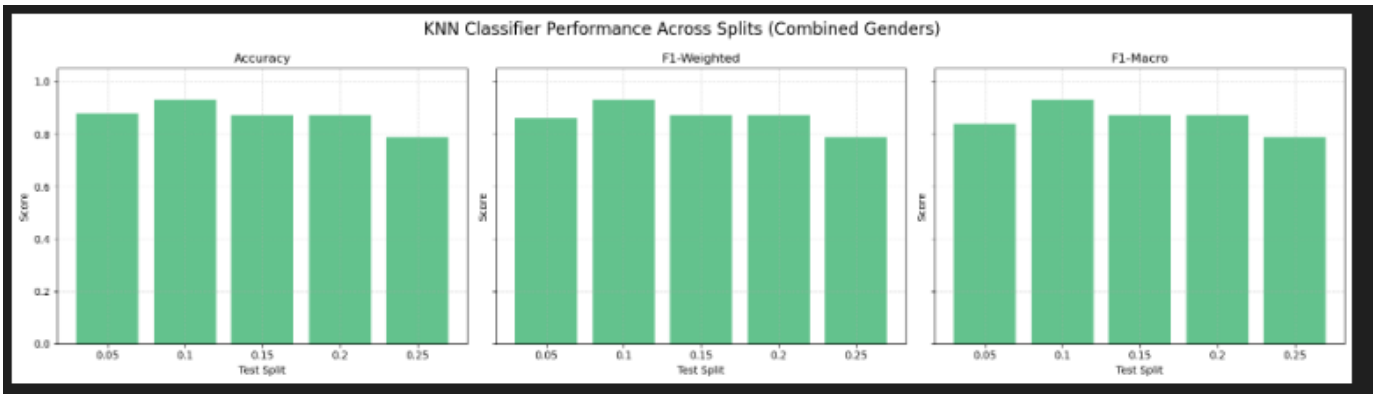
(b) Split 10%



(c) Split 20%



(d) Split 25%

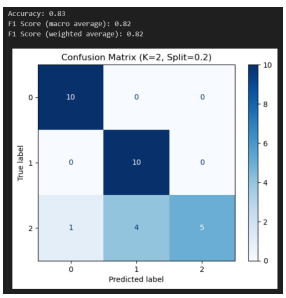


(e) Classifier Performance Across Test Splits

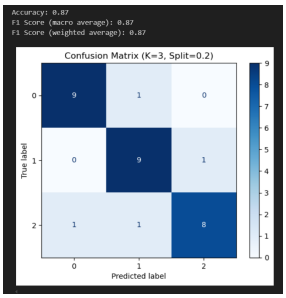
Figure 1: Confusion Matrices for Different Test Splits

1.2 Varying K

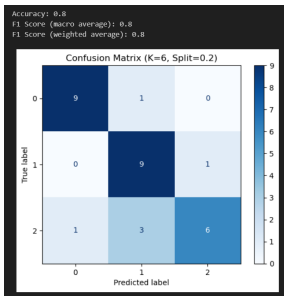
Classifiers performance was lowest at k=2, as can be seen in the increased confusion matrix in class 2 predictions, but improved significantly at k=3, yielding best F1 and accuracy scores. Performances remained stable between k=3 and k=5, with with consistent classification across all vowel classes, whilst larger k values (6, 7) introducing mild underfitting, and k=10 mantaining solid performance. k=3 will be used in subsequent experiments, since it has best overall balance between precision and generalization.



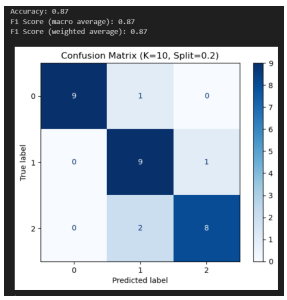
(a) K = 2



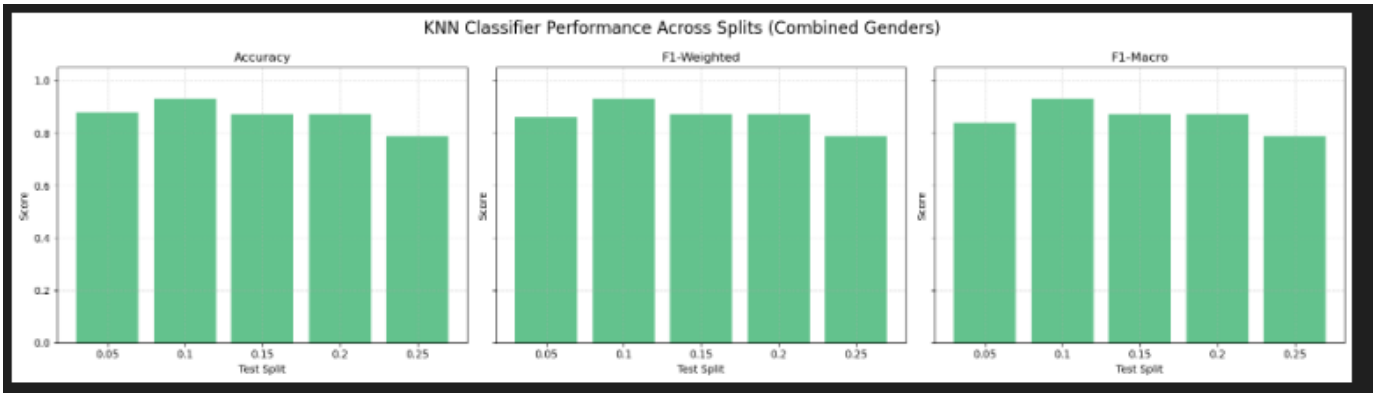
(b) K = 3



(c) K = 6



(d) K = 10

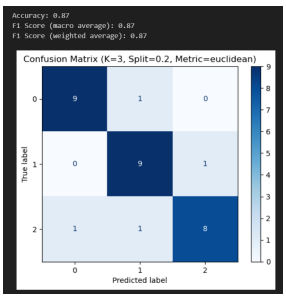


(e) Classifier Performance Across K Values

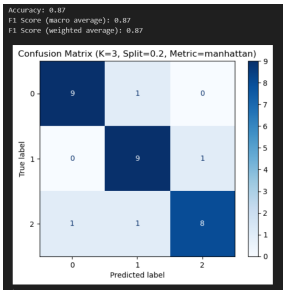
Figure 2: Confusion Matrices for Different K Values

1.3 Distance Metrics

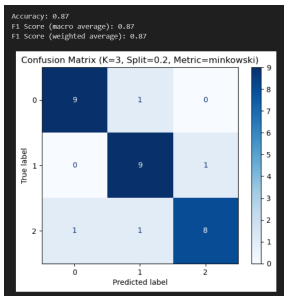
The Euclidean, Manhattan, and Minkowski distance metrics all produced identical and reliable results with an F1 score of 0.87, showing minimal confusion. In contrast, the Hamming metric performed extremely poorly, achieving only 0.33 accuracy and misclassifying every instance as the same class. Future experiments will use Euclidean.



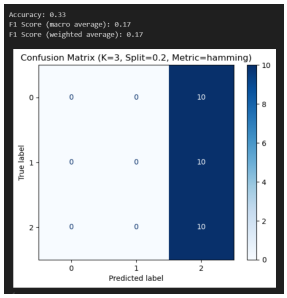
(a) Euclidean



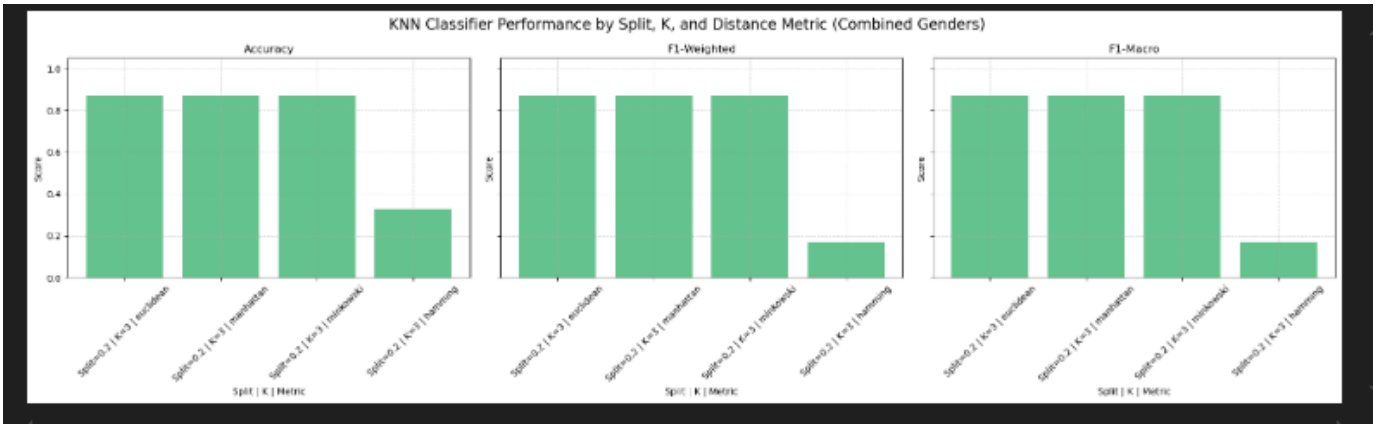
(b) Manhattan



(c) Minkowski



(d) Hamming



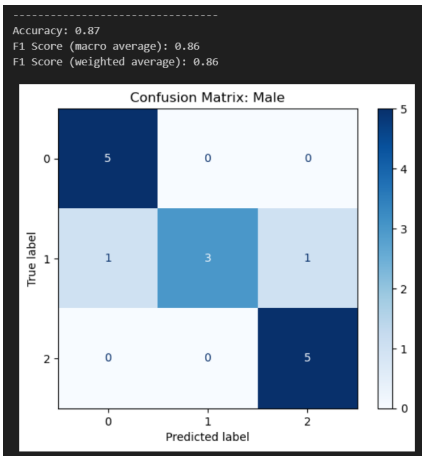
(e) Classifier Performance Across Distance Metrics

Figure 3: Confusion Matrices for Different Distance Metrics

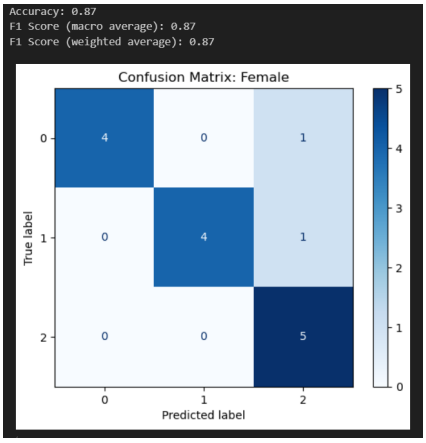
1.4 Gender Segregation and Combination

Classifiers trained on male and female data separately, both achieved high accuracy of 0.87, though male classifier showed a slight bit more confusion in class 1 classification and an F1 score of 0.86 when compared to females 0.87. Female model had minor misclassification between 0 and 2 classes, but maintained a clean diagonal dominance. When both genders were combined performance remained equally as strong, with more balanced classification. This indicated that the mixed-gender data generalizes just as well, if not better than gender specific models.

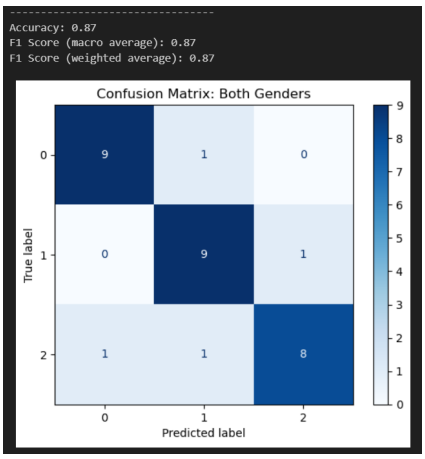
The best F1 Score and Accuracy achieved by the parameter-tuned model, was an F1 Score of 0.87 and Accuracy of 87%.



(a) Male

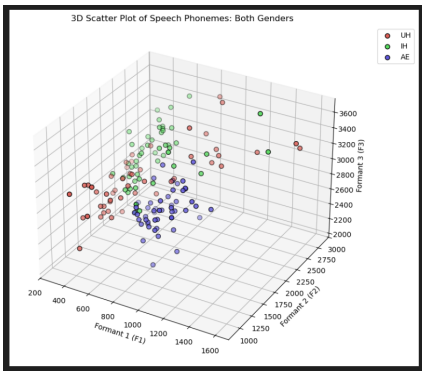


(b) Female

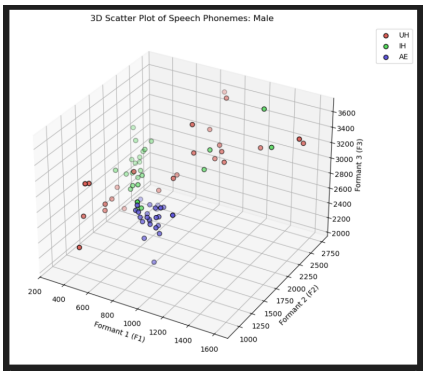


(c) Both Genders

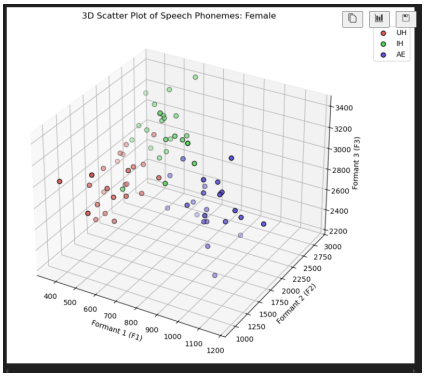
Scatter plots show clear clustering of all three phonemes. AE occupies lower F1 and midrange F2/F3, IH occupies mid-F1 and higher F2, and UH at higher F1 values. While male and female exhibit similar overall separations, female data points are slightly more clustered and compact, suggesting less variability in formant values among female speakers.



(a) Both Genders



(b) Male



(c) Female

Figure 5: Confusion Matrices for Gender Separations

1.5 Vowel Phoneme Confusion

Across all confusion matrices above, most fequent off-diagonal errors happen between second and third phoneme classes, i.e. IH and AE. These two vowel sounds are sometimes swapped, indicating ovelapping formant values more than UH.

2 Generative AI Usage

Generative AI [1] was only used to fix any errors that came along the way, and to help with formant frequency value understanding, making sure they align with their respective arpabet symbol.

References

[1] OpenAI, "ChatGPT (GPT-4.5)," Accessed: May 2025. [Online]. Available: https://chat.openai.com/

[2] F. Pedregosa *et al.*, "Scikit-learn: Machine Learning in Python," *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011. [Online]. Available: https://scikit-learn.org/stable/modules/neighbors.html.