Vowel formants predict place of articulation of following consonant in Kannada: Evidence for the autosegmental model of phonology

Praveen N. Pallegar

Introduction

In the field of human language comprehension, we see an enormous computational task in the ability of humans to perceive, parse, and comprehend large amounts of information in real-time. On the semantic and syntactic levels of analysis, a great deal of research has shown that humans lighten this computational load by virtue of semantic/syntactic frameworks and association networks that allow us to leverage the predictive capacity of language to our advantage [2,6]. However, relatively less research has been done on the predictive capacity of phonemes, the smallest distinguishable unit of sound.

A landmark paper in phonology proposed the autosegmental model for phonology, claiming, among other ideas, that feature parsing of phonemes is not a bijective function, and that each phoneme additionally carries information about its environment [4]. Thus, such secondary information could be used to simultaneously predict phonetic environments, thereby further reducing the computational load of phonetic comprehension. More recently, Gow et al. showed that listeners could differentiate between segments that had merged completely through place assimilation. The authors hypothesized that the preceding vowel retained information about the underlying forms, allowing listeners to disambiguate meaning [3].

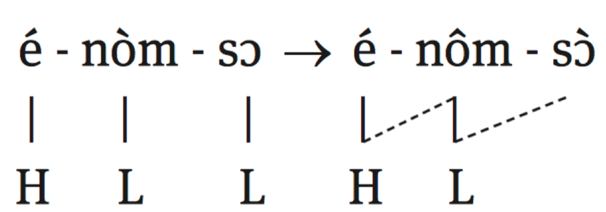


Fig 1. Tonal assimilation in terms of autosegmental theory [5]

In the present study, I use machine learning techniques to show that vowels have a significant predictive capacity over the following consonant's place of articulation.

Methods

Data Collection

Waveform recordings were collected from two native speakers of Kannada and vowels preceding retroflex and non-retroflex consonants were isolated. Resonant frequencies of these vowels (known as formants) were extracted and used as features. Each example was labeled as 1 (if preceding a retroflex consonant) or 0 (if preceding a non-retroflex consonant). Below is an example of such extraction using Praat software.

ನನಗೆ ಸ್ವಲ್ಪ ಹಾಲು ಕೊಡಿ → [nɐnəgɛ.sʷɐlpə.hä:lu.koɖi]

Methods

Data Collection

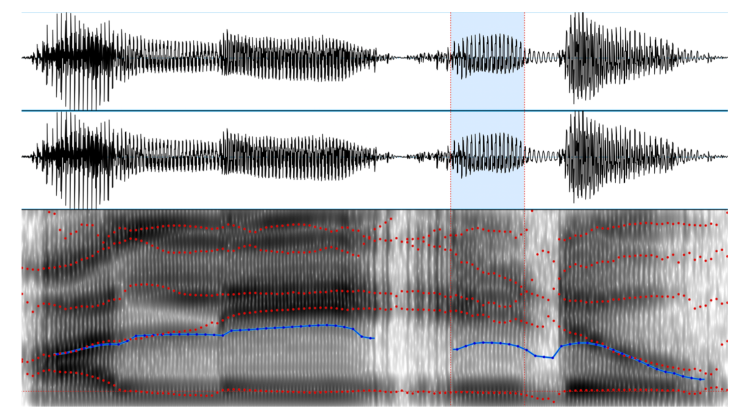


Fig 2. Feature extraction preceding retroflex consonant

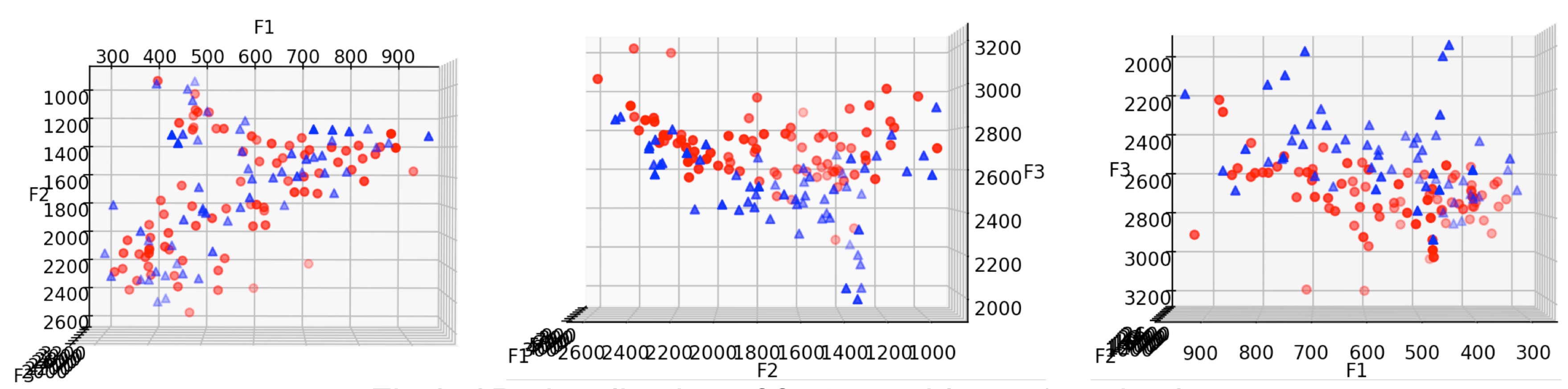
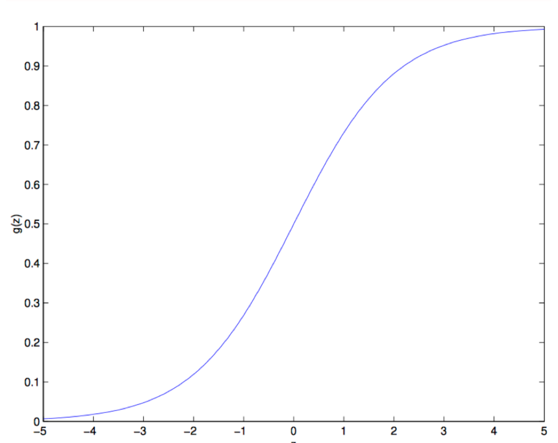


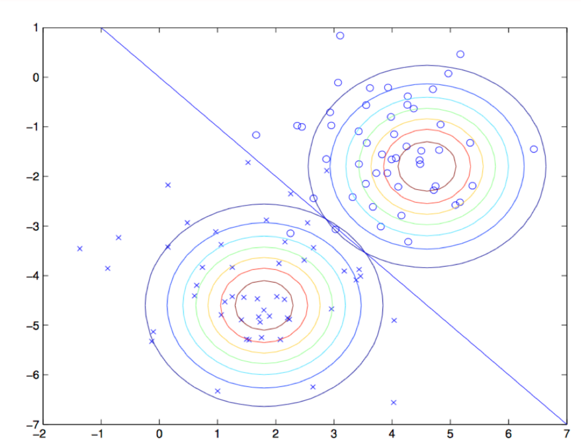
Fig 3. 3D visualization of features; blue = 1, red = 0

Machine Learning Algorithms

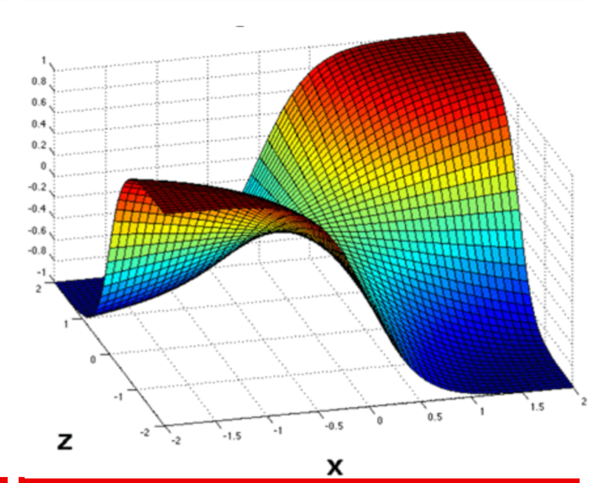
Logistic Regression



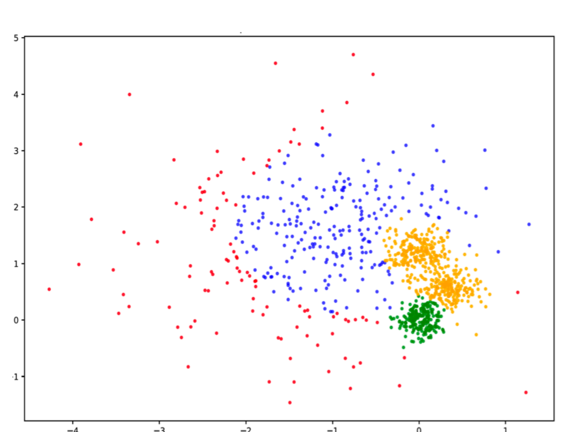
Gaussian Discriminant Analysis



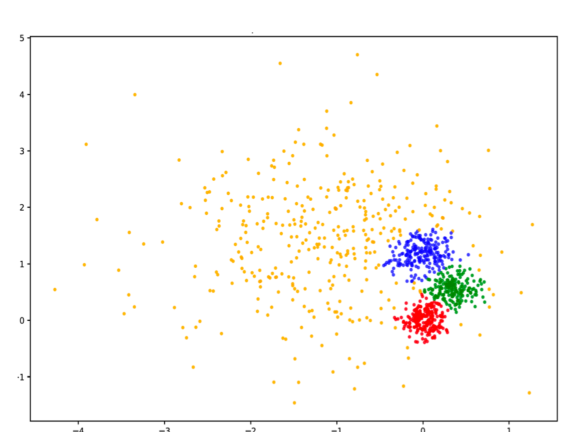
Support Vector Machine



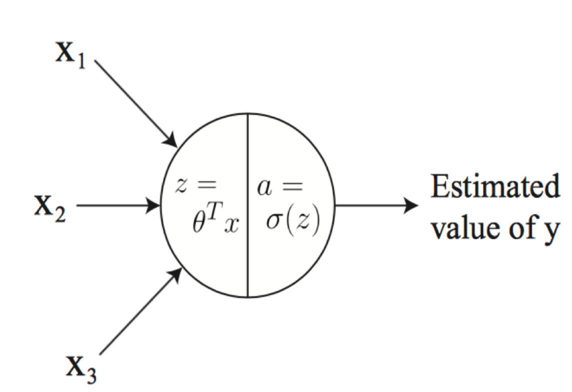
Unsupervised GMM



Semi-supervised GMM



Neural Network



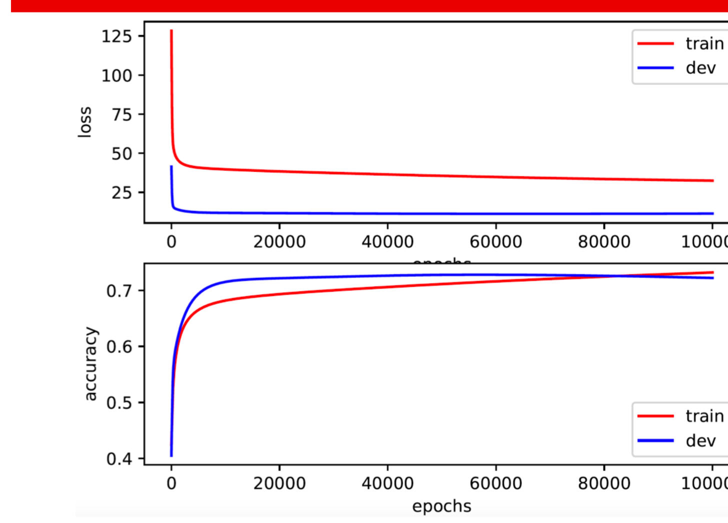
Testing and Optimization

- K-fold cross-validation (k = 5) was used to measure accuracy on Gaussian Discriminant Analysis and Logistic Regression models
- Backwards feature selection was used to optimize models and measure predictive capacity of each feature

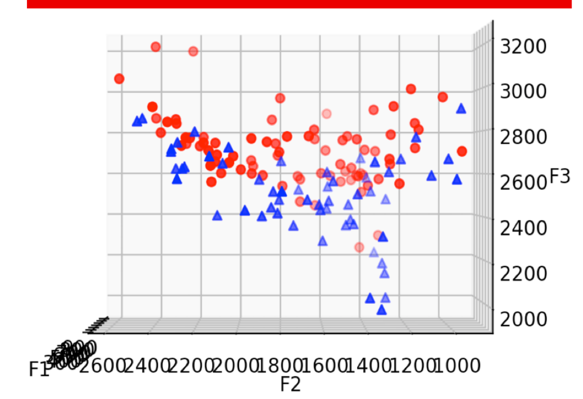
Results

Model	Accuracy
Logistic Regression	0.727
GDA	0.734
SVM (sigmoid kernel)	0.741
Unsupervised GMM	0.529
Semi-supervised GMM	0.714
Neural Network	0.614

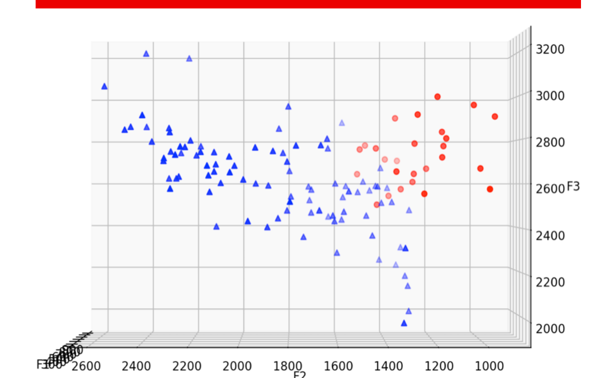
Neural Network



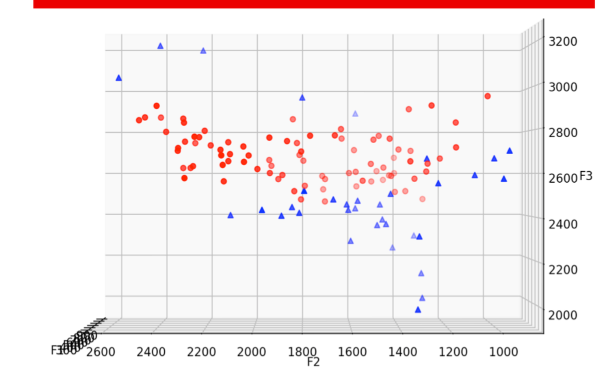
Data



Unsupervised GMM



Semi-supervised GMM



Optimization

- Feature selection on Logistic Regression and GDA had highest accuracy when all features where evaluated though F3 had a much higher predictive capacity than F1 and F2
- Sigmoid kernel performed best compared to RBF, polynomial, and linear kernels
- Neural network performed best with 1 hidden layer of size 15 and a learning rate of 0.001 (sigmoid function was the activation of both hidden and output layers)

Discussion

Though the SVM approach had the best test accuracy, I believe that deep learning could perform equally as well, if not better, given more data. In tuning the model, adding more neurons or increasing the learning rate from the optimum caused the model to drastically overfit, but this could be prevented given a larger dataset. I have shown that Kannada vowels have a significant predictive capacity over their environment, providing support for the autosegmental theory of phonology.

References

[1] Carrington, A. M., Fieguth, P. W., & Chen, H. H. (2014). A new Mercer sigmoid kernel for clinical data classification. *2014 36th Annual International Conference of the IEEE Engineering in Medicine and Biology Society*. doi: 10.1109/embc.2014.6945092

[2] Freunberger, D., & Roehm, D. (2016). Semantic prediction in language comprehension: evidence from brain potentials. *Language, Cognition and Neuroscience*, *31*(9), 1193–1205. doi: 10.1080/23273798.2016.1205202

[3] Gow, D. W. (2003). Feature parsing: Feature cue mapping in spoken word recognition. *Perception & Psychophysics*, *65*(4), 575–590. doi: 10.3758/bf03194584

[4] Goldsmith, John. 1976a. *Autosegmental phonology*. Cambridge, MA: Massachusetts Institute of Technology dissertation.

[5] Hyman, L. M. (2014). How autosegmental is phonology? *The Linguistic Review*, *31*(2). doi: 10.1515/tlr-2014-0004

[6] Staub, A., & Clifton, C. (2006). Syntactic prediction in language comprehension: Evidence from either...or. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *32*(2), 425–436. doi: 10.1037/0278-7393.32.2.425