Finding vowels without phonology?

Kathleen Currie Hall and E. Allyn Smith
The Ohio State University

Introduction

1. Linguists generally identify contrastive sounds in a language by identifying (near) minimal pairs (e.g. “to” and “tea” indicate that [u] and [i] are different phonemes), as well as relying on alternations and other consistent patterns to identify sound relationships.

2. Segmental inventories are defined and the features needed to describe each segment are derived from such examinations of a particular language.

3. Problems: how many pairs do you need to establish contrast? how close to a minimal pair do they have to be? e.g. does the minimal difference between “riding” [raiɾɪŋ] and “writing” [rʌiɾɪŋ], or between “Cyclops” [saiklaps] and “micron” [mʌikran] mean that [ai] and [ʌi] are separate vowels in Canadian English (Mielke, Armstrong, & Hume 2003)? how do you tell if any given utterance contains one vowel or another?

4. There are no clear answers to such questions, which require extensive research into the phonological system.

5. In addition to phonological information about segments, though, we also have access to phonetic information. We can turn to the native speakers of a language to see whether their productions fall into consistent clusters that can be identified.

6. Peterson & Barney (1952) introduced the concept of correlating acoustic measurements with phonemic categories.

7. To identify the salient clusters, however, they used their knowledge of what words were intended by each speaker to encircle the points that belong to each vowel: “the closed loops for each vowel have been drawn arbitrarily to enclose most of the points” (182).

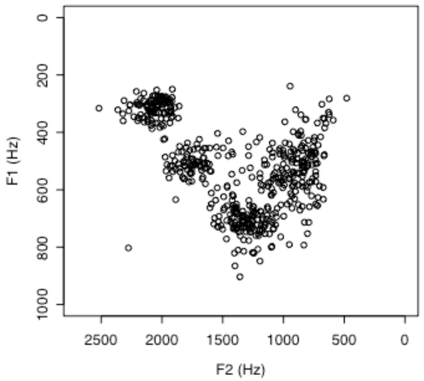
8. Our proposal: use modern statistical techniques to automatically identify particular clusters in the acoustic data. This provides an insight into what speakers consistently produce as potentially distinct segments.

An Example: Clustering Greek Vowels

9. A single native male speaker of Athenian Greek was recorded reading a Samarin list of 200 words commonly found in languages.

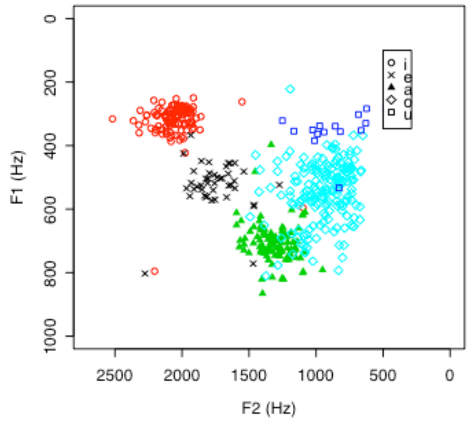
10. F1 and F2 measurements were made at the midpoint of each vowel that naturally occurred in the resulting words – a representative sample of the vowels in Greek.

11. Taking just the raw acoustic data and making an F1/F2 plot of the vowel space results in the following graph:



12. In eyeballing this graph, it seems clear that there are four – or maybe five – clusters.

13. The actual labels for these vowels (as determined by relying on a larger phonological and phonetic examination of Greek, in addition to extant literature (e.g. Botinis 1981, Jongman et al. 1989, Fourakis et al. 1999) and native speaker intuitions) are as follows – there are five vowels in Greek:



14. Can a computer find these? K-Means Clustering (MacQueen 1967) is a way of dividing a data set into K ‘clusters’ (disjoint subsets of the data points) without having to label the clusters according to a variable. It works roughly as follows:

a. The user specifies K (i.e. K=3, etc)
b. The K-Means algorithm (given below) initially distributes the data points at random into the K clusters.[1]
c. The centre point of each cluster is computed (this point is essentially the one whose values are the means over the cluster points for each variable; here, F1 and F2).
d. The Sum of Squares algorithm (a part of the larger K-means algorithm) basically computes the distance between each point and its cluster centre and then computes the distance between the point and the other cluster centres.
e. The K-Means algorithm then determines which cluster centre is the minimum distance from the point in question, and, in the event that that centre is not the one to which the point is currently assigned, the algorithm reassigns the point to the new cluster whose center it is closest to.

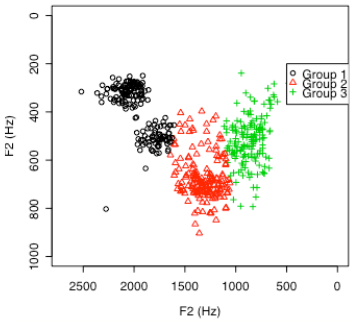
---

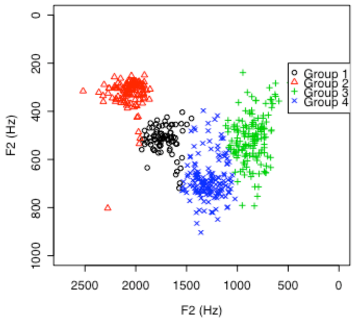
[1] Because the initial cluster membership is assigned at random, it is important to run this algorithm several times for each K value to verify that it produces essentially the same clusters each time.

f. After points have been reassigned, the mean values of the variables within each cluster have changed, and so they are re-computed.
g. Once the cluster centre values have been re-computed, however, it is necessary to determine whether there are any points that are now closer to centres in other clusters than than their own centre. Thus, step (e) is repeated.
h. These last two steps repeat until no points are reassigned, at which point the algorithm terminates.

$$J = \sum_{j=1}^{K} \sum_{n \in S_j} |x_n - \mu_j|^2,$$

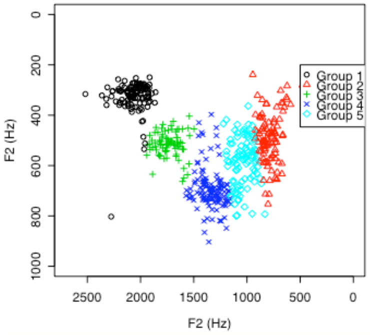
15. Here are results of 3 / 4 / 5 / 6 clusters on the Greek vowel data:[2]

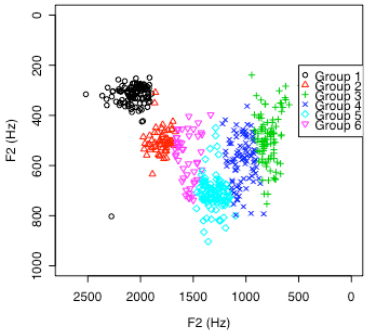




---

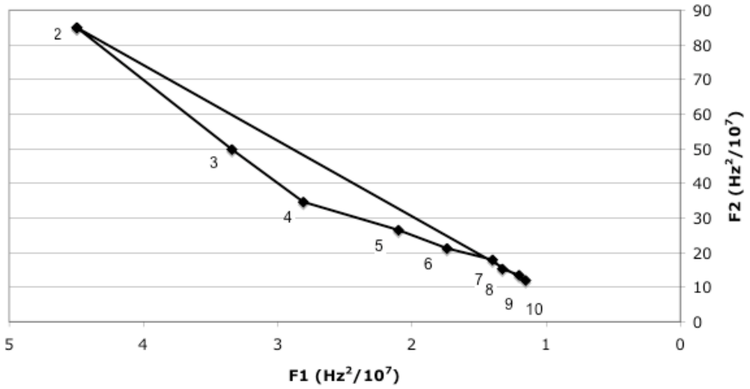
[2] This K-means clustering was computed using R (Venables et al. 2002). If you would like to create such graphs yourself, please e-mail us for our R code.





16. The five cluster graph above does not mirror the actual labels in (11); the clustering algorithm has a slight tendency to split along the front/back dimension more than it should to accurately mark the vowel groups.

17. So far, this just tells you for each number you specify, how they could be clustered. How do we determine which of these clusterings minimizes variance (and thus best represents the vowel space)? Plot the final Sum of Squares results for each K once the algorithm has terminated (i.e. sum of squared distances from all of the final means for each K):



18. The values in the graph above were divided by 10^7 merely for improved readability.

19. Once the variance for each K is plotted, one method of determining which clustering is optimal is to look for an ‘elbow’ in the graph (the technical term!). Here, we see one at K=4, but the standard line is provided as a measure for cases when the elbow is less clear.[3]

20. Based on an elbow at K=4, K-Means Clustering has predicted that Greek has 4 vowels. Looking at the graph for K=4, we see that the algorithm correctly clustered [i], [e], [a], and [o], but assimilated the [u]’s to the [o] cluster.

Some Improvements

21. The low frequency of [u], combined with its proximity to the very frequent and dispersed [o], makes this vowel in Greek particularly difficult to “find” by algorithm. Perhaps a larger sample would be more effective, especially if a higher cluster density were attained.

22. Modelling the vowels in terms of Hertz reflects their acoustic rather than their auditory properties. Mapping the perceptual space (in ERB, mels, Bark, etc.) might prove to be more fruitful.

23. Multidimensional clustering could also improve accuracy: including F3, duration, etc. as other measurements in computing clusters.

24. Depending on the goal of clustering (e.g., to determine whether [ai] and [ʌi] are separate vowels in Canadian English, given the rest of the segmental inventory), it may be worthwhile to allow knowledge about other phonemes in the language to inform the clustering algorithm.

25. For example, knowing that a particular back vowel is preceded by a coronal could indicate that it would be produced with a more front vowel than would otherwise be expected.

26. This sort of co-articulation can be normalized in K-means clustering by adding or subtracting from the numeric Hertz values. E.g. Stevens & House (1963)

---

[3] The use of a standard line or other technique for calculating the elbow is an area of statistical methods that requires further study.

found that [u]'s preceded by a coronal had an average F2 that was 350 Hz higher than those in neutral contexts. By subtracting 350 Hz from e.g. all vowels that have "low F1" and "low F2," and that were preceded by a coronal, the effect of the coronal can be subtracted out and make the actual [u] cluster more dense, hence easier for the K-means algorithm to locate.

27. Similarly, knowledge about stressed/unstressed vowels, part of speech, etc., might shed light on how to adjust clusters for a particular purpose (see, e.g., Moon & Lindblom 1984).

28. Classification and regression trees allow us to build some of these factors directly into the model instead of adjusting measurements.

Future Directions

29. Applying these techniques to more speakers – comparing results across speakers could allow the actual clusters in the language to emerge.

30. Applying these techniques to other languages – different statistical approaches may be favored based on examining different situations.

31. Applying these techniques to phonological and sociolinguistic problems like Canadian Raising – clustering may indicate how many distinct, consistent phonetic vowel categories exist in a language, or where the category boundaries are if the number of categories has already been established.

32. Exploring the impact on theories of acquisition – can kids acquire abstract phonetic categories through this mass of exemplars, and use this knowledge to inform their phonological inferences?

33. Using this technique in field research – clustering could be a way to get an overview of the phonetic categories in a language before extensive phonological research can be done.

34. Mining extant recordings – corpora and collections of field tapes can be used as sources of clustering data, both to explore the statistical techniques that would be most fruitful and as a practical means of gaining more descriptive information from data that might otherwise go unstudied.

References

Botinis, A. (1981). The Greek vowels, *Working Papers in Linguistics - Phonetics* (Vol. 21, pp. 29-36): Lund University.

Breiman, L., Friedman, J. H., Olshen, R. A., & Stone, C. J. (1984). Classification and regression trees. Belmont, CA: Wadsworth & Brooks.

Fourakis, M. S., Botinis, A., & Kalsaiti, M. (1999). Acoustic characteristics of Greek vowels. *Phonetica, 56*, 28-43.

Hastie, T., Tibshirani, R., & Friedman, J. H. (2001). *The elements of statistical learning: data mining, inference, and prediction*. New York: Springer.

Hawks, J. W., & Fourakis, M. S. (1995). The perceptual vowel spaces of American English and Modern Greek: a comparison. *Language and Speech, 38*(3), 237-252.

Jongman, A., Fourakis, M. S., & Sereno, J. A. (1989). The acoustic vowel space of Modern Greek and German. *Language and Speech, 32*(3), 221-248.

MacQueen, J. B. (1967). Some methods for classification and analysis of multivariate observations. *Proceedings of the 5th Berkeley Symposium on Mathematical Statistics and Probability*, 1, 281-297.

Mielke, J., Armstrong, M., & Hume, E. (2003). Looking through opacity. *Theoretical linguistics, 29*, 123-139.

Miller, J. D. (1989). Auditory-perceptual interpretation of the vowel. *Journal of the Acoustic Society of America, 85*(5), 2114-2134.

Moon, S.-J., & Lindblom, B. (1994). Interaction between duration, context, and speaking style in English stressed vowels. *Journal of the Acoustic Society of America, 96*(1), 40-55.

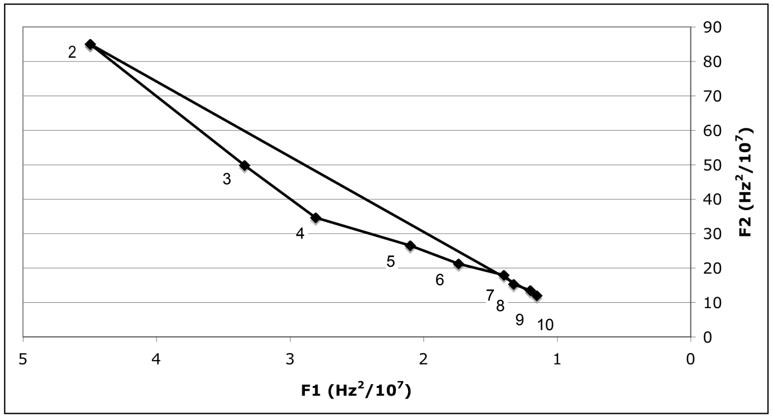
Peterson, G. E., & Barney, H. L. (1952). Control methods used in a study of vowels. *Journal of the Acoustical Society of America, 24*(2), 175-184.

Stevens, K. N., & House, A. S. (1963). Perturbations of vowel articulations by consonantal context: an acoustical study. *Journal of Speech and Hearing Research, 6*(2), 111-128.

Venables, W. N., & Ripley, B. D. (1999). *Modern applied statistics with S-PLUS* (3rd ed.). New York: Springer.

Venables, W. N., Smith, D. N., and the R Development Core Team. (2002). *An Introduction to R*. Ms.

Weisstein, E. W. (2006) K-means clustering algorithm. From *MathWorld*, available at http://mathworld.wolfram.com/K-MeansClusteringAlgorithm.html. Accessed Feb. 2006.


2
3
4
5
6
7
8
9
10
90
80
70
60
50
40
30
20
10
0
F2 (Hz^2/10^7)
5
4
3
2
1
0
F1 (Hz^2/10^7)