Comparing k-means and OPTICS clustering algorithms for identifying vowel categories

Background: In descriptive phonetics, clustering algorithms are often used to evaluate the dispersion of vowel categories (e.g. De Boer & Kuhl 2003; Czoska et al 2015; Shi et al 2019). The *k*-means algorithm (Forgy 1965) is most commonly used for this purpose but has major limitations: i) it requires a specified number of clusters, ii) it assumes that clusters are spheres of similar size (Han et al. 2011), and iii) can be skewed by outliers. OPTICS is an algorithm that uses areas of high density in the data to identify clusters in contrast to k-means, which partitions all points in the data into *k* clusters (Ankerst et al 1999). As a result, OPTICS avoids many of the pitfalls of *k*-means: it organically assigns the number of clusters, can learn clusters of varying shape and density, and separates prototypical members of a cluster from outliers.

The aim of this study is to explore whether OPTICS is a viable alternative to *k*-means as a clustering algorithm for learning patterns in vowel data. We compare the performance of OPTICS and *k*-means on two English datasets and find that when given noisier data, OPTICS has several advantages over *k*-means for characterizing vowel spaces.

Methodology: In this study, we selected two datasets of English monophthongs to represent highly-controlled 'lab' speech (Hillenbrand et al.1995) and less-controlled 'corpus' speech (Buckeye Corpus: Pitt et al. 2005). We tested *k*-means and OPTICS clustering with each dataset, using speaker-normalized midpoint formant measures as the parameters.

Results: In lab speech, both *k*-means and OPTICS clustering successfully identified clusters corresponding to labeled vowel categories. In the corpus dataset (Fig. 1 on the following page), the identified clusters roughly correspond to labeled vowel categories. Both algorithms identify an 'extra' cluster based on either area (*k*-means, cluster 1) or density (OPTICS, cluster 1). *k*-means partitions the entire space into roughly evenly sized clusters, which results in inaccurate boundaries (e.g. between /u/ and /ɪ/). On the other hand, OPTICS identifies accurate vowel centers, but doesn't attempt to categorize every observation. Because of these different approaches to clustering, OPTICS is more well-suited to illustrating the core vowel space.

Conclusions: Both *k*-means and OPTICS are able to approximate superficial structure in the vowel space, but differ in their underlying mathematical approach. Unlike *k*-means, density-sensitive methods like OPTICS can identify areas of high density even in noisy data and extract likely vowel centers from that structure. This study emphasizes the importance of choosing an appropriate clustering algorithm and highlights the potential of density-based algorithms like OPTICS for analyses of vowel spaces.

References:

Ankerst, M., Breunig, M., Kriegel, H-P., and Sander, J. (1999). OPTICS: ordering points to identify the clustering structure. *ACM SIGMOD*, 28(2) (1999): 49-60. **Czoska, A., Katarzyna K., & Karpinski, M.** (2015). Polish infant directed vs. adult directed speech: Selected acoustic-phonetic differences. *18th International Congress of Phonetic Sciences,* Glasgow, UK. **De Boer, B., & Kuhl, P. K.** (2003). Investigating the role of infant-directed speech with a computer model. *Acoustics Research Letters Online*, 4(4). **Forgy, E W.** (1965). Cluster analysis of multivariate data: efficiency versus interpretability of classifications. *Biometrics*. 21 (3). **Han, J., Pei, J., & Kamber, M.** (2011). *Data mining: concepts and techniques*. Elsevier. **Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K.** (1995). Acoustic characteristics of American English vowels. *The Journal of the Acoustical Society of America*, 97(5). **Pitt, M. A., Johnson, K., Hume, E., Kiesling, S., & Raymond, W.** (2005). The Buckeye corpus of conversational speech: Labeling conventions and a test of transcriber reliability. *Speech Communication*, 45(1). **Shi, Y., Renwick M., & Maier, F.** (2019). Improved vowel labeling for prenasal merger using customized forced alignment. Poster presented at the *178th Meeting of the Acoustical Society of America*, San Diego, CA.

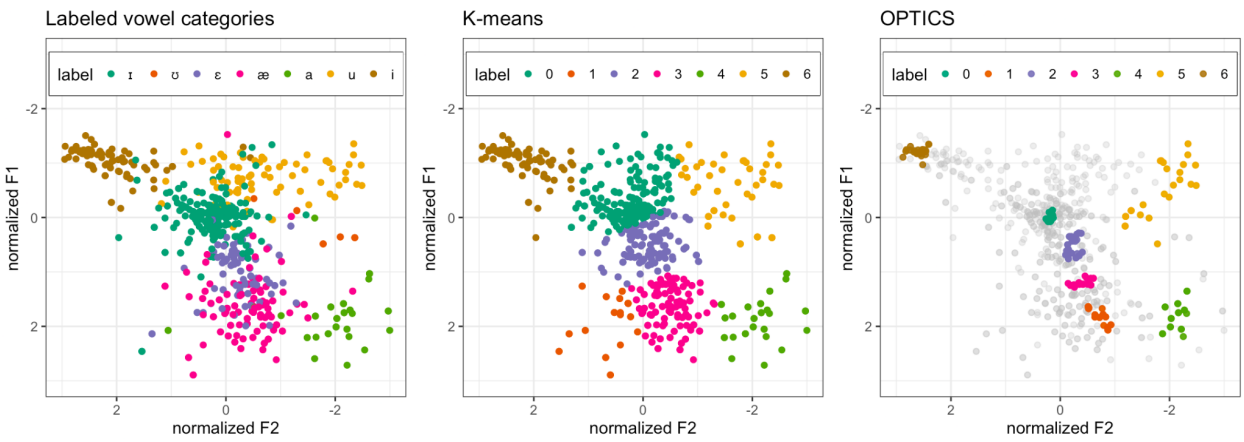

Figure 1: Comparing labeled vowel categories (left), k-means clustering (center), and OPTICS clustering (right)