2021. *Proc Ling Soc Amer* 6(1). 403–410. https://doi.org/10.3765/plsa.v6i1.4977.

Automatic phonetic classification of vocalic allophones in Tol

Marie Bissell*

Abstract. The aim of the present study involving automatic phonetic classification of /e/ and /u/ tokens in Tol is two-fold: first, I test existing claims about allophonic variation within these vowel classes, and second, I investigate allophonic variation within these vowel classes that has yet to be documented. The acoustic phonetic classifications derived in the present study contribute to a more detailed understanding of the allophonic systems operating within the Tol language. Operationalizing machine learning algorithms to investigate under-resourced, indigenous languages has the potential to provide detailed insights into the acoustic phonetic dynamics of a diverse range of vocalic systems.

Keywords. machine learning; under-resourced languages; vowels; allophones

1. Introduction. The use of machine learning algorithms for linguistic research has gained traction more broadly in recent years, including in the subfields of semantics (Liang and Potts 2015, Potts 2019, Boleda 2020), phonology (Linzen 2019, Pater 2019, Rawski and Heinz 2019), and dialectology (Hartley 2005, Evanini 2008). In the realm of phonetics, most work involving machine learning algorithms has focused on constructing and testing automatic speech recognition programs (Sagayama 1989; Deng and Li 2013; Agarwalla and Sarma 2016; Ault, Perez, Kimble, and Wang 2018); however, some recent studies have harnessed machine learning algorithms for the purpose of investigating acoustic aspects of phonemic contrast (Jones, Meakins, and Muawiyath 2012; Renwick and Ladd 2016; Renwick and Nadeu 2019). The current study extends these recent acoustic inquiries to examining applications of machine learning algorithms to describing the acoustic characteristics of allophonic systems operating within an under-resourced, endangered language spoken in Central America. Although machine learning technologies have been applied to data from endangered languages primarily for purposes of automatic speech recognition in the past (Besacier, Barnard, Karpov, and Schultz 2014; Rey and Nagy 2018; Mohammed 2020), the current study aims to use a machine learning algorithm to generate clusters of acoustic similarity for vocalic productions to explore patterns of phonologically-conditioned allophonic splits in Tol.

Tol, a Hokan language spoken by around 500 indigenous Tolupan people living on a reservation in south-central Honduras near Tegucigalpa, has been impressionistically described by several researchers, including Fleming and Dennis (1977) and Holt (1999). Dennis and Dennis (1983) put together a Spanish-Tol dictionary, while the latter two works described the sound systems of the Tol language in more depth. Although these descriptions were quite detailed, both were fundamentally impressionistic in nature. The current study set out to examine two particular claims about vocalic allophones that appeared in both works using acoustic data from the a large-scale corpus designed for research on phonetic typology, *Vox Clamantis* (Salesky, Chodroff, Pimentel, Wiesner, Cotterell, Black, and Eisner 2020).

Fleming and Dennis (1977) wrote in considerable depth about allophones of the /e/ and /u/ vowel classes in Tol. They described /e/ as being produced as [e] preceding /ŋ/ and as [ɛ] preceding other syllabic codas. They also wrote that /u/ was sometimes produced as [ʊ] preceding /s/ or

[*] I would like to acknowledge the Tol speakers whose acoustic data was analyzed in this project. Author: Marie Bissell, The Ohio State University (bissell.43@osu.edu).

/c^h/ and as [u] preceding other syllabic codas. The current study aimed to analyze acoustic data from a corpus of Tol speech (Salesky et al. 2020) to adjudicate these two allophonic claims.

2. Methods. The corpus used in this study was composed of speech samples from spoken bible readings by Tol speakers that were gathered as part of the *Vox Clamantis* project (Salesky et al. 2020). The vowel tokens analyzed in the current study had been previously segmented by Salesky et al. (2020). I operationalized k-means clustering (Huang 1998, Steinley 2006, Kaufman and Rousseeuw 2009), a machine learning algorithm that locates a pre-specified number of clusters in a data set, to automatically cluster pre-coda tokens of /e/ (n = 27,402) and /u/ (n = 18,983) from these recordings in order to determine whether allophonic splits were identifiable via acoustic measurements. For this analysis, I measured the first two formants of each vowel token at 25%, 50%, and 75% duration: all six of these measurements per vowel were ultimately submitted to the kmeans() function, such that the algorithm had a reasonable amount of acoustic data to work with from several timepoints throughout each vowel token.

2.1. K-MEANS CLUSTERING ALGORITHM. I implemented this machine learning algorithm in R software (R Core Team 2020) with the kmeans() function from the *stats* package. Before running the kmeans() function for the measurements I had made for each vowel class, I first went about calculating the appropriate number of clusters for the algorithm to look for in the data set for each vowel class. Running the k-means clustering algorithm requires the user to manually select the number of clusters for the algorithm to look for in the data, so computing the optimal number of clusters for each vowel class occurred first chronologically. Although several methods for identifying an optimal number of clusters exist (Pham, Dimov, and Nguyen 2005; Chiang and Mirkin 2010; Kodinariya and Makwana 2013), I selected the silhouette method due to its relative popularity and ease of implementation in R software.

I used the silhouette() function from the *cluster* package in R (Maechler, Rousseeuw, Struyf, Hubert and Hornik 2019) to calculate a silhouette coefficient for each set of vowel data. A silhouette coefficient is fundamentally a measure of how similar each data point is to other data points in its own cluster versus to data points in other clusters (Rousseeuw 1987). This coefficient ranges from -1 to +1, with higher coefficient values corresponding to points being more similar (i.e., having lower Euclidean distances) to other points in their own cluster. To calculate this value for each vowel's acoustic measurements, the silhouette() function completes these steps:

(1) Compute the average distance of a given point *i* to all other points in point *i*'s cluster. [a(*i*)]
(2) Compute the average distance of that given point *i* to all other points in the nearest neighboring cluster. [b(*i*)]
(3) Compute the silhouette coefficient by calculating (b(*i*) – a(*i*))/max(b(*i*), a(*i*)).
(4) Repeat steps 1 through 3 for every point in the data set, then average all of the silhouette coefficients to arrive at the overall silhouette coefficient for that data set.

After computing a series of silhouette coefficients for various numbers of clusters per vowel class data set, I then selected the number of clusters that corresponded to the highest positive silhouette coefficient. More detailed information about this process appears later in the results section, where these comparisons are represented in several visualizations.

Once I had located the optimal silhouette coefficient for each of the two vowel classes, I ran a k-means algorithm in R with the kmeans() function in the *stats* package. This function took two arguments: the six acoustic measurements associated with each vowel and a numeric value for k, which is the number of clusters associated with the optimal silhouette coefficient from those analyses described previously. The k-means algorithm completes these steps:

(1) Randomly choose centroids for *k* number of clusters.
(2) Calculate the distance from each data point to each randomly chosen centroid.
(3) Assign each data point to the closest centroid (i.e., cluster) in terms of Euclidean distance.
(4) Calculate a new centroid for each cluster by averaging the mean locations of all points assigned to that cluster.
(5) Repeat steps 2 through 4 until the cluster centroids stop moving.

3. Results. First, I calculated the appropriate number of clusters for each of the two sets of vowel tokens using the silhouette method. Figures 1 and 2 show the results of these silhouette analyses for /e/ and /u/ tokens, respectively. Both silhouette method results indicated that two clusters was the optimal value for *k* in the k-means algorithm, suggesting that this number of clusters maximized each data point's similarity to those other data points assigned to its own cluster.

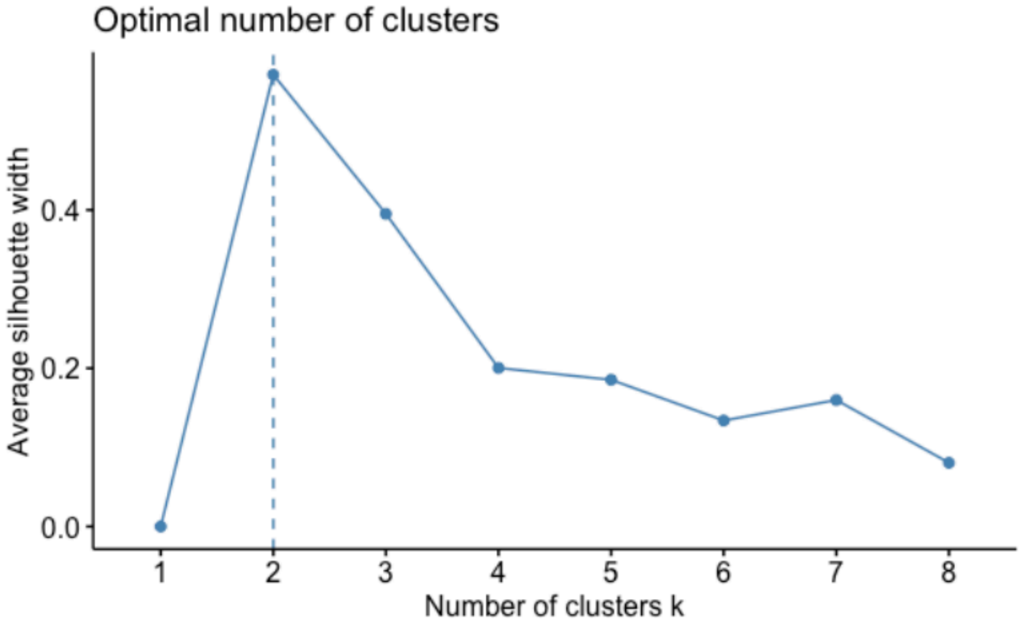

Figure 1. Optimal number of clusters for the k-means algorithm for the /e/ tokens as determined by the silhouette method.

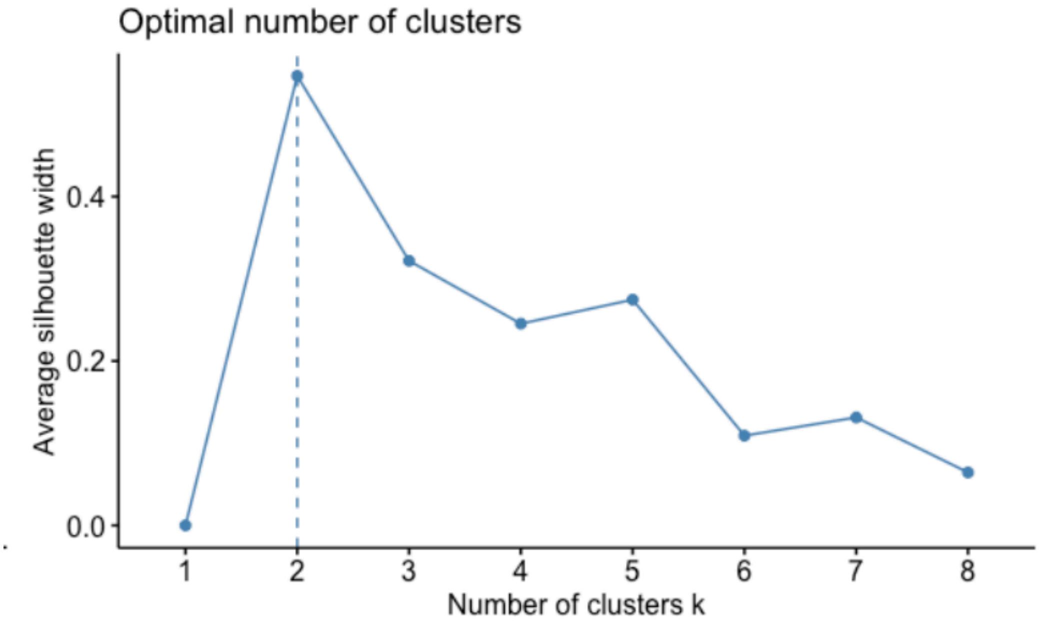

Figure 2. Optimal number of clusters for the k-means algorithm for the /u/ tokens as determined by the silhouette method.

The results of the k-means algorithms, both of which were set to locate two clusters as per the results of the silhouette method analysis, are shown in Figures 3 and 4: for the sake of plot readability, the dimensions shown on the axes are mean F1 and F2 values per following environment at 25% duration.

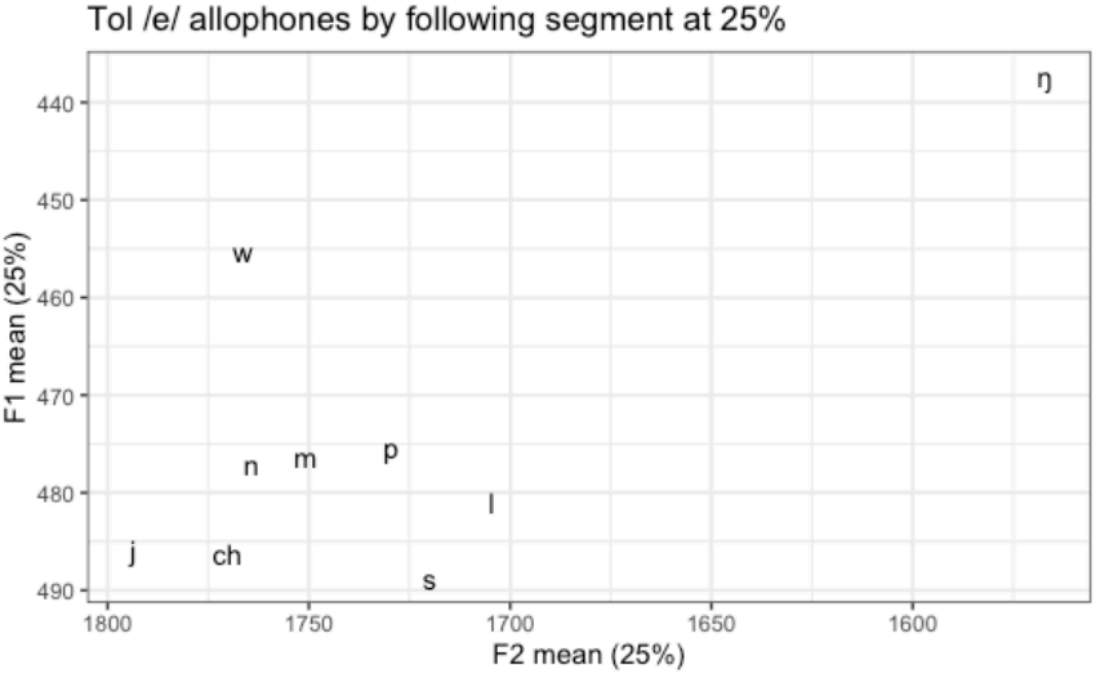

Figure 3. Mean F1 and F2 values for /e/ vowel tokens by following environment.

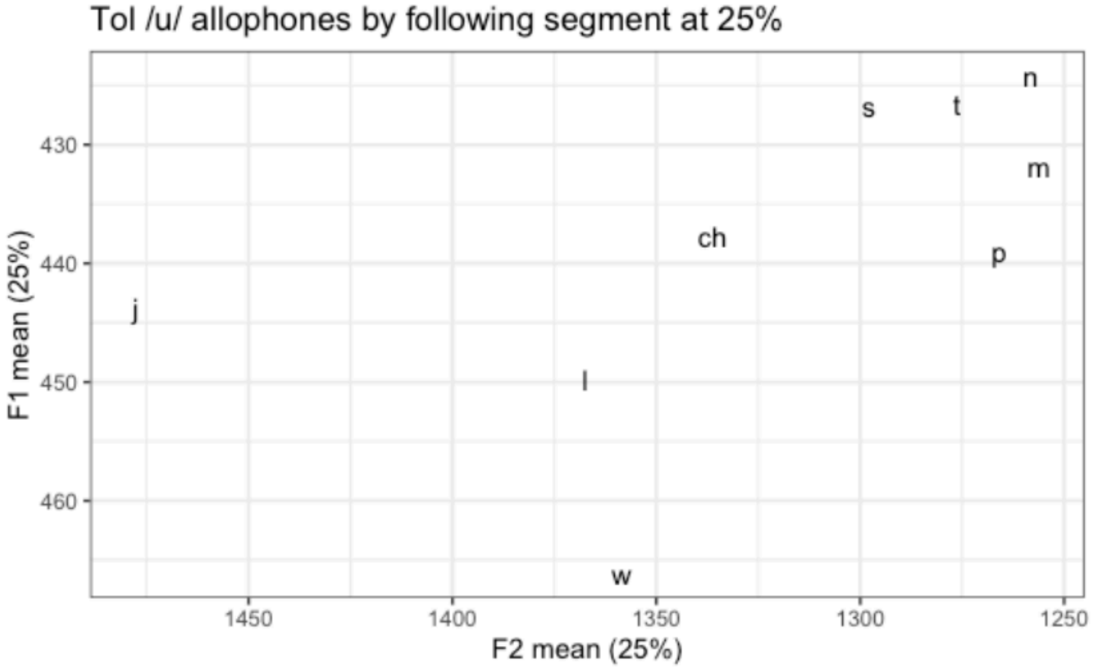

Figure 4. Mean F1 and F2 values for /u/ vowel tokens by following environment.

For the /e/ tokens, pre-/ŋ/ productions were distinct from productions in other following environments both in terms of F1 and F2 at 25% duration. This finding matches the impressionistic reports offered by Fleming and Dennis (1977) and Holt (1999). The results of the current study provide supplementary acoustic evidence in favor of this allophonic distinction within the /e/ class in the Tol language.

For the /u/ tokens, pre-/j/ productions were distinct from productions in other following environments primarily in terms of F2 at 25% duration. The frontness of these pre-/j/ tokens suggests that they are more [ʊ]-like than the tokens in other following environments. While Fleming and Dennis (1977) and Holt (1999) impressionistically observed [ʊ]-like productions for the /u/ vowel class in pre-/s/ and pre-/c^h/ environments, the current study shows acoustic evidence to support that [ʊ]-like productions are most common in pre-/j/ environments.

4. Discussion and conclusions. The findings reported in the current study partially support and partially challenge previous accounts of vocalic allophony in the /e/ and /u/ vowel classes in Tol. My analysis of /e/ tokens supported existing impressionistic descriptions of /e/ allophony in Tol, but my analysis of /u/ tokens showed that pre-/j/ productions were consistently closer to [ʊ] than pre-/s/ or pre-/c^h/ productions were.

For the /u/ productions in pre-/j/ environments, it is possible that there is a phonological motivation having to do with natural classes at work: what complicates this analysis is that there appears to be one allophone in pre-/j/ environments and another allophone in pre-/w/ environments. One possible explanation for the [ʊ]-like allophone appearing only before /j/ is coarticulation, but it is not yet clear how the /uj/ sequence specifically differs from the /uw/ sequence in Tol such that one would trigger an allophonic change due to coarticulation and the other would not. Another possibility is that /j/, which has been previously described by Fleming and Dennis (1977) as a syllabic coda that can occur after /u/, is actually functioning as a sort of vowel-glide sequence whose internal structure is distinct in some important way from the internal structure of vowel-consonant sequences. The current analysis does not claim to adjudicate among these possibilities due to lack of relevant data at this time, but it is certainly the case that there are several plausible motivations for this allophonic split.

The primary aim of the current study was to examine whether acoustic evidence could be located to support or dispute impressionistic descriptions of vocalic allophony in the Tol language for the /e/ and /u/ vowel classes. My results for /e/ support existing descriptions of /e/ allophony but my results for /u/ did not support existing descriptions of /u/ allophony. Because Tol is an under-resourced language with limited acoustic documentation, the current study offers a more detailed perspective on the acoustic operations of its allophonic systems in vowels: operationalizing machine learning techniques to investigate the acoustic dynamics of under-studied vocalic systems has the capacity to expand knowledge about allophony and its triggers cross-linguistically. In particular, k-means clustering is a very useful tool for exploring allophonic patterns in acoustic space, and in the future this kind of machine learning algorithm has potential to be used for a wider variety of clustering tasks, including phoneme identification and tone systems.

References

Agarwalla, Swapna & Kandarpa Kumar Sarma. 2016. Machine learning based sample extraction for automatic speech recognition using dialectal Assamese speech. *Neural Networks* 78. 97–111. https://doi.org/10.1016/j.neunet.2015.12.010.

Ault, Shaun V., Rene J. Perez, Chloe A. Kimble & Jin Wang. 2018. On speech recognition algorithms. *International Journal of Machine Learning and Computing* 8(6). 518–523. https://doi.org/10.18178/ijmlc.2018.8.6.739.

Besacier, Laurent, Etienne Barnard, Alexey Karpov & Tanja Schultz. 2014. Automatic speech recognition for under-resourced languages: A survey. *Speech Communication* 56. 85–100. https://doi.org/10.1016/j.specom.2013.07.008.

Boleda, Gemma. 2020. Distributional semantics and linguistic theory. *Annual Review of Linguistics* 6. 213–234. https://doi.org/10.1146/annurev-linguistics-011619-030303.

Chiang, Mark Ming-Tso & Boris Mirkin. 2010. Intelligent choice of the number of clusters in k-means clustering: an experimental study with different cluster spreads. *Journal of Classification* 27(1). 3–40. https://doi.org/10.1007/s00357-010-9049-5.

Deng, Li & Xiao Li. 2013. Machine learning paradigms for speech recognition: An overview. *IEEE Transactions on Audio, Speech, and Language Processing* 21(5). 1060–1089. https://doi.org/10.1109/TASL.2013.2244083.

Evanini, Keelan. 2008. Classifying and clustering dialects of North American English. *North East Student Colloquium on Artificial Intelligence (NESCAI).*

Fleming, Ilah & Ronald K. Dennis. 1977. Tol (Jicaque): Phonology. *International Journal of American Linguistics* 43(2). 121–127. https://doi.org/10.1086/465467.

Hartley, Laura C. 2005. The consequences of conflicting stereotypes: Bostonian perceptions of US dialects. *American Speech* 80(4). 388–405. https://doi.org/10.1215/00031283-80-4-388.

Holt, Dennis. 1999. Tol (Jicaque). Munich: Lincom Europa.

Huang, Zhexue. (1998). Extensions to the k-means algorithm for clustering large data sets with categorical values. *Data Mining and Knowledge Discovery* 2(3). 283–304. https://doi.org/10.1023/A:1009769707641.

Jones, Caroline, Felicity Meakins & Shujau Muawiyath. 2012. Learning vowel categories from maternal speech in Gurindji Kriol. *Language Learning* 62(4). 1052–1078. https://doi.org/10.1111/j.1467-9922.2012.00725.x.

Kaufman, Leonard & Peter J. Rousseeuw. 2009. *Finding groups in data: An introduction to cluster analysis*. Hoboken, NJ: John Wiley & Sons.

Kodinariya, Trupti M. & Prashant R. Makwana. 2013. Review on determining number of clusters in K-Means Clustering. *International Journal* 1(6). 90–95.

Liang, Percy & Christopher Potts. 2015. Bringing machine learning and compositional semantics together. *Annual Review of Linguistics* 1(1). 355–376. https://doi.org/10.1146/annurev-linguist-030514-125312.

Linzen, Tal. 2019. What can linguistics and deep learning contribute to each other? Response to Pater. *Language* 95(1). e99–e108. http://doi.org/10.1353/lan.2019.0015.

Maechler, Martin, Peter Rousseauw, Anja Struyf, Mia Hubert & Kurt Hornik. 2013. cluster: Cluster Analysis Basics and Extensions. R package version 1.14.4.

Mohammed, Siraj. 2020. Using machine learning to build POS tagger for under-resourced language: the case of Somali. *International Journal of Information Technology* 12. 717–729. https://doi.org/10.1007/s41870-020-00480-2.

Pater, Joe. (2019). Generative linguistics and neural networks at 60: Foundation, friction, and fusion. *Language* 95(1). e41–e74. https://doi.org/10.1353/lan.2019.0009.

Pham, Duc Truong, Stefan S. Dimov & Chi D. Nguyen. 2005. Selection of K in K-means clustering. *Proceedings of the Institution of Mechanical Engineers, Part C: Journal of Mechanical Engineering Science* 219(1). 103–119. https://doi.org/10.1243/095440605X8298.

Potts, Christopher. 2019. A case for deep learning in semantics: Response to Pater. *Language* 95(1). e115–e124. https://doi.org/10.1353/lan.2019.0019.

R Core Team. 2020. R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Rawski, Jonathan & Jeffrey Heinz. 2019. No free lunch in linguistics or machine learning: Response to Pater. *Language* 95(1). e125–e135. http://doi.org/10.1353/lan.2019.0021.

Renwick, Margaret E. L., & D. Robert Ladd. 2016. Phonetic distinctiveness vs. lexical contrastiveness in non-robust phonemic contrasts. *Laboratory Phonology: Journal of the Association for Laboratory Phonology* 7(1). http://doi.org/10.5334/labphon.17.

Renwick, Margaret E. L. & Marianna Nadeu. 2019. A survey of phonological mid vowel intuitions in central Catalan. *Language and Speech* 62(1). 164–204. https://doi.org/10.1177/0023830917749275.

Rey, Lyndon & Naomi Nagy. 2018. Automatic documentation of Faetar's [i]: A methodology for ddiscovering vowel space using artificial neural networks. *Géolinguistique* 18. https://doi.org/10.4000/geolinguistique.306.

Rousseeuw, Peter J. 1987. Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics* 20. 53–65. https://doi.org/10.1016/0377-0427(87)90125-7.

Sagayama, Shigeki. 1989. Phoneme environment clustering for speech recognition. *International Conference on Acoustics, Speech, and Signal Processing* 1. 397–400. https://doi.org/10.1109/ICASSP.1989.266449.

Salesky, Elizabeth, Eleanor Chodroff, Tiago Pimentel, Matthew Wiesner, Alan W. Black & Jason Eisner. 2020. A large-scale corpus for phonetic typology. *Proceedings of the 58th Annual Meeting for the Association of Computational Linguistics.*

Steinley, Douglas. 2006. K-means clustering: a half-century synthesis. *British Journal of Mathematical and Statistical Psychology* 59(1). 1–34. https://doi.org/10.1348/000711005X48266.