Towards a measure of optimization in natural vowel systems

INTRODUCTION

Studies looking at the emergence and evolution of phonological systems have shown that, given sufficient evolutionary time, organizations of the articulatory space emerge in which phonemes are maximally distinctive (e.g. Steels, 1997; de Boer, 2000; Oudeyer, 2005; de Boer & Zuidema, 2010). However, there has been little investigation into the typological description of articulatory optimization across the world's languages. It is not known, for example, how optimized natural vowel-spaces actually are, or whether the vowels of, for example, English are more or less distinctive than those of, for example, Swahili. Here, I introduce a methodology for measuring exactly this. The aims of this research were to demonstrate (a) that such a measure is possible, and (b) that this measure could have practical uses within Evolutionary Linguistics.

METHODS

To develop a method for measuring vowel-space optimization, it is necessary to have a representation of how the vowels of a language relate to each other within the space formed by the oral cavity. Peterson and Barney (1952) showed that when the first and second formant frequencies of a set of vowel sounds are plotted on reversed logarithmic axes, a visualization of the vowel-space emerges. Plots of this kind thus provide a convenient way to observe the spatial relationships between a language's vowel sounds.

Data collection

Global distribution of languages in the sample.

Since formant frequency data for the world's languages are not readily available, the data had to be collected manually. Audio recordings were downloaded from the UCLA Phonetics Lab Archive (Ladefoged & Blankenship, 2007) for acoustic analysis. A sample of 70 languages was selected at random, and the Praat software application (Boersma & Weenink, 2011) was used to extract the formant frequency data for each vowel in each of the 70 languages for a total of 415 vowels.

Transformation to a psychoacoustic scale

The vowels' formant frequencies were transformed to a psychoacoustic scale. This is necessary because the human auditory system works logarithmically, such that high frequency sounds appear closer together than low frequency sounds. The mel scale (Stevens, Volkmann, & Newman, 1937) was chosen simply because its calculation is the least computationally expensive; other scales (Bark, etc.) give the same final results.

Measuring the distance between vowels

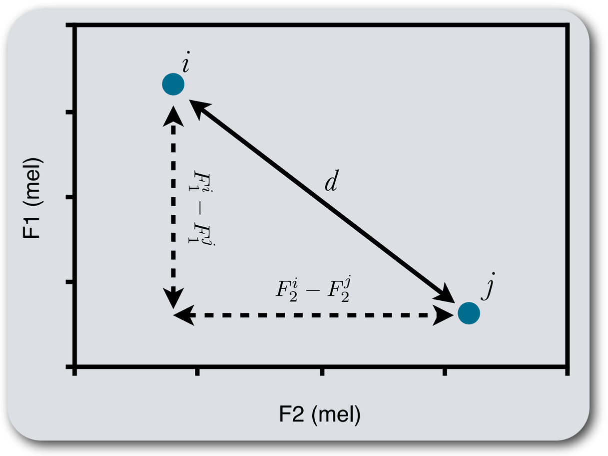


Diagram showing the calculation of the Euclidean distance *d* between vowels *i* and *j*.

The Euclidean distance in mels between any pair of vowels *i* and *j* can be calculated by using the formant values as Cartesian coordinates in a two-dimensional space. Thus

$$d = \sqrt{(F_1^i - F_1^j)^2 + (F_2^i - F_2^j)^2}$$

gives the distance d between vowel i (F_1^i, F_2^i) and vowel j (F_1^j, F_2^j). Since the third formant (F_3) is also important in the perception of vowels, the vowel-space was extended into three dimensions with F_3 plotted on the z-axis. The distance between vowels was therefore measured across a three-dimensional space, capturing the three most salient vowel formants.

Measuring optimization

Following Liljencrants and Lindblom (1972), the inverse-square law from theoretical physics is used to get a sense of how optimally distributed the vowels are within the vowel-space. To illustrate, take the analogy of two particles with equal electrical charge. The particles repel each other with a force that is inversely proportional to the square of the distance between them. If the particles are confined to a limited space, they will move apart until their mutual distance is maximized. The optimal state is the one in which the potential energy in the system is minimized, which is given by the sum of the inverse of the square of the distance between each pair of particles. Assuming that a similar process occurs in the evolution of vowel systems (in the sense that vowel systems seek to maximize perceptual contrast), the potential energy in a vowel system can be calculated thus:

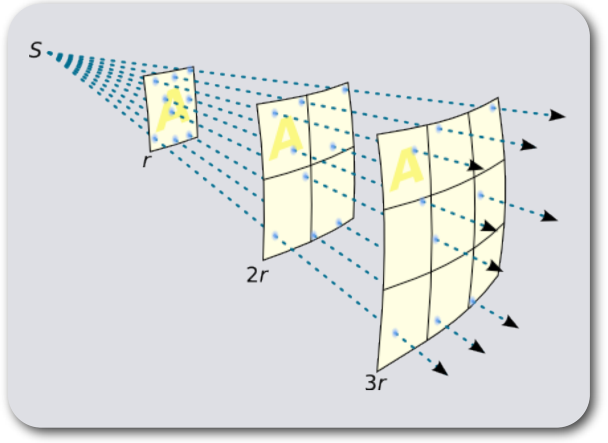


Illustration of the inverse-square law. The force of repulsion decreases with the square of the distance from the source.

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{\sum_{f=1}^{3} (F_f^i - F_f^j)^2}$$

where energy E is the sum of the inverse of the square of the Euclidean distance between vowels i and j for every possible pairing of vowels (n being the total number of vowels in the system). When the total energy in a vowel system is minimized, the optimization of the vowel system is maximized, and vice versa. Thus, by taking the inverse of E, we derive a number that corresponds to vowel-space optimization.

Standardization of the measure

To standardize this measure across individuals, and test for statistical significance, Monte Carlo techniques were used. For each language, we generate 100,000 randomized vowel-sets. We calculate the total energy present in each randomized vowel-set, take the inverse, and then calculate a standard score by comparing the natural vowel-set against the mean and standard deviation of the randomized ones. A z-score greater than 0 suggests the vowels are further apart than one would expect by chance.

Summary

This method captures two key properties of an optimized system: effectiveness and order. The inverse-square law tells us how effective (i.e. how perceptually distinctive) the distribution of vowels is given the finite space in which they exist. The application of Monte Carlo techniques tells us how ordered (i.e. how non-random) the vowel system is by comparing the natural system against ones which are known to be stochastic.

RESULTS

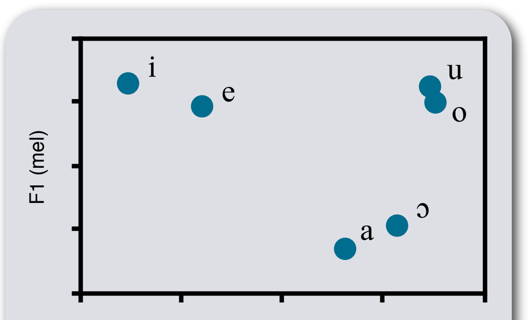
There is a high degree of variation in vowel-space optimization, which ranges from -0.969 for the Azerbaijani language to 6.144 for the Nyangumarta language (mean = 2.087, SD = 1.556). See the plots below for three examples.

Optimization inversely correlates with vowel inventory size (r = -0.514, n = 70, p < 0.000006), such that languages with large inventories of vowels tend to be less optimized. This should be obvious, since as we add more vowels into a system, the energy forcing them apart will increase.

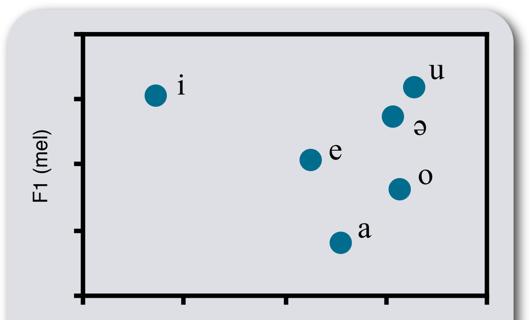
Hay and Bauer (2007) showed that a positive correlation exists between vowel inventory size and speaker population size. This is also true of the sample studied here (r = 0.454, n = 70, p < 0.00008). I wanted to test whether there also exists a correlation between vowel-space optimization and speaker population size (controlling for inventory size), which could have interesting implications for the evolution of phonological systems. However, the results suggest that there is no such correlation (r = 0.089, n = 70, p = 0.459). (Population data was taken from Ethnologue (Lewis, 2009).)

Although population size shows no significant interaction with vowel-space optimization, it illustrates the kinds of patterns that may be identified with more in-depth analyses.

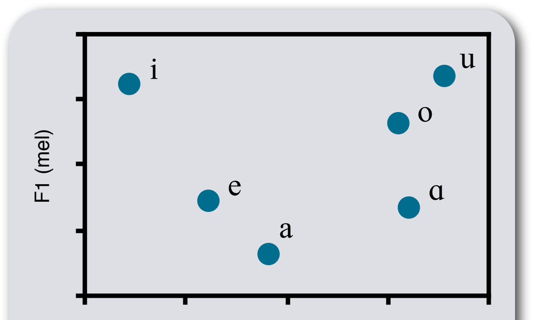
The vowel-spaces of 3 languages which represent examples of an unoptimized system, a typical system, and an optimized system. Note that the optimization of the vowel-space is reflected in the z-score.



Defaka (Niger-Congo): z-score = 0.073



Itelmen (Chukotko-Kamchatkan): z = 2.348



Farsi (Indo-European): z-score = 3.768

DISCUSSION

This research has demonstrated that it is possible to construct a measure of vowel-space optimization – which to my knowledge has not been attempted previously. Such a measure could have practical uses in several areas of Linguistics. However, more work will be required to make more robust conclusions. In particular, these 70 languages represent just 1 or 2 per cent of global linguistic diversity.

Strengths

The optimization score does seem to intuitively fit with what unoptimized and optimized vowel-spaces ought to look like, and I therefore very tentatively suggest that this method offers a reliable measure of vowel-space optimization.

This may be useful in at least two areas of enquiry: firstly, the results show that natural vowel-spaces tend towards optimized organizations, which is implicitly assumed by simulations of the emergence of phonemic systems (e.g. Steels, 1997; de Boer, 2000; Oudeyer, 2005; de Boer & Zuidema, 2010). Secondly, the results may be useful in typological studies (e.g. Ember & Ember, 2007; Hay & Bauer, 2007; Lupyan & Dale, 2010) by allowing us to see whether vowel-space optimization interacts with external social properties.

Challenges

Firstly, the raw data used in this study are inherently fuzzy – speakers do not consistently produce vowels with precisely the same formant frequencies, which could amount to significant inaccuracies that skew the results. Secondly, the simulations assume *a priori* that the vowels of the natural language delineate the maximum space available for a given speaker. This could introduce ceiling effects such that the randomized languages may only make use of a subset of the full articulatory space that is realistically possible. Thirdly, it is currently difficult to make comparisons between languages with different inventory sizes because the score is not inventory-size neutral.

Main conclusions

This paper has demonstrated a method for measuring the level of optimization in the perceptual vowel-spaces of natural languages, which could be of use to future linguistic research. Although further work will be required to improve on this method, it could have practical implications for understanding how vowel systems evolve in order to adapt to changing environmental, social, and cognitive demands.

REFERENCES

Boersma, P., & Weenink, D. (2011). Praat: Doing phonetics by computer (Version 5.3) [Software]. Available from www.praat.org

de Boer, B. (2000). Self-organization in vowel systems. *Journal of Phonetics*, *28*, 441–465.

de Boer, B., & Zuidema, W. (2010). Multi-agent simulations of the evolution of combinatorial phonology. *Adaptive Behavior*, *18*, 141–154.

Ember, C. R., & Ember, M. (2007). Climate, econiche, and sexuality: Influences on sonority in language. *American Anthropologist*, *109*, 180–185.

Hay, J., & Bauer, L. (2007). Phoneme inventory size and population size. *Language*, *83*, 388–400.

Ladefoged, P., & Blankenship, B. (2007). *The UCLA Phonetics Lab Archive*. Los Angeles: UCLA Department of Linguistics. Available from archive.phonetics.ucla.edu

Lewis, M. P. (Ed.) (2009). *Ethnologue: Languages of the world* (16th ed.). Dallas, TX: SIL International. Available from www.ethnologue.com

Liljencrants, J., & Lindblom, B. (1972). Numerical simulation of vowel quality systems: The role of perceptual contrast. *Language*, *48*, 839–862.

Lupyan, G., & Dale, R. (2010). Language structure is partly determined by social structure. *PLoS ONE*, *5*(1), e8559.

Oudeyer, P.-Y. (2005). The self-organization of speech sounds. *Journal of Theoretical Biology*, *233*, 435–449.

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

Steels, L. (1997). The synthetic modeling of language origins. *Evolution of Communication*, *1*, 1–34.

Stevens, S. S., Volkmann J., & Newman, E. (1937). A scale for the measurement of the psychological magnitude pitch. *Journal of the Acoustical Society of America*, *8*, 185–190.

This research was supported by a Carnegie-Cameron Taught Postgraduate Bursary. Many thanks are due to Jim Hurford and the anonymous second marker of the essay that ignited this research.

Jon William Carr

Language Evolution and Computation Research Unit
School of Philosophy, Psychology and Language Sciences
University of Edinburgh

j.w.carr@sms.ed.ac.uk

THE UNIVERSITY OF EDINBURGH