Exploring the role of production in predicting vowel inventories[*]

Nathan Sanders
Williams College
nsanders@williams.edu

Jaye Padgett
UC Santa Cruz
padgett@ucsc.edu

1 Overview

Since the seminal work of Liljencrants and Lindblom (1972), a key testing ground for functional, evolutionary, or emergentist approaches to sound systems has been the typology of vowel inventories (for example, Lindblom 1986, Schwartz et al. 1997a, de Boer 2000).

An important innovation of Schwartz et al.'s Dispersion-Focalization Theory (DFT) was calculating the optimality ("energy") of a vowel system as a weighted combination of *two* separate auditory parameters:

(1) **dispersion**: maximization of the auditory distance between vowels (as in Liljencrants and Lindblom 1972)

focalization: maximization of the importance of "focal" vowels such as [i] and [y].

DFT makes reasonably good predictions, matching or approximating many of the attested vowel systems found in the UCLA Phonological Segment Inventory Database (UPSID; Maddieson 1984, Maddieson and Precoda 1989).

However, there are still numerous attested vowel systems that DFT does not predict to be optimal, most notably, many systems containing [ə] and other more centrally located vowels, which happen to be less articulatorily extreme than more peripheral or focal vowels.

We argue that an articulatory parameter should be added to DFT, and we report promising preliminary results from modifications to DFT which model articulatory effort.

2 How Dispersion-Focalization Theory works

In DFT, vowel systems are compared according to their total "energy" according to the distribution and types of vowels in each system. The lower a vowel system's total energy is, the more optimal it is.

For a given vowel system $\{\mathrm{V}_1, \ldots, \mathrm{V}_N\}$, each vowel V_i is characterized by its first four formants $\langle \mathrm{F1}_i, \mathrm{F2}_i, \mathrm{F3}_i, \mathrm{F4}_i \rangle$, measured in Bark.[1]

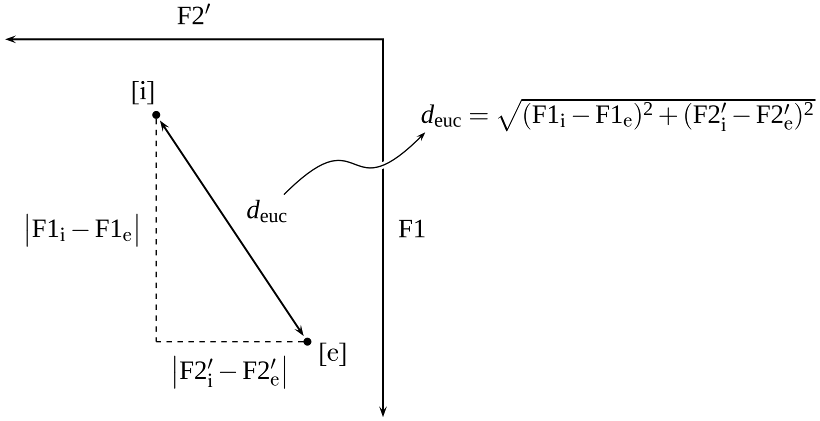
A system's total energy E_{DF} (2) is just the simple sum of its dispersion energy E_D (§2.1) and its focalization energy E_F (§2.2):

(2) $E_{DF} = E_D + E_F$

2.1 Dispersion

The dispersion energy of a vowel system is a measure of the overall auditory distance between the vowels in the system. The basic auditory distance between two vowels V_i and V_j is the euclidean distance d_{euc} between them in the auditory space based on their values of F1 and *effective* F2, a.k.a. "F2 prime", a hypothesized perceptual integration of F2, F3, and F4, symbolized as F2′ (Carlson et al. 1970, 1975):

(3)


F2′
[i]
$d_{\mathrm{euc}} = \sqrt{(\mathrm{F1_i} - \mathrm{F1_e})^2 + (\mathrm{F2'_i} - \mathrm{F2'_e})^2}$
d_{euc}
$|\mathrm{F1_i} - \mathrm{F1_e}|$
F1
[e]
$|\mathrm{F2'_i} - \mathrm{F2'_e}|$

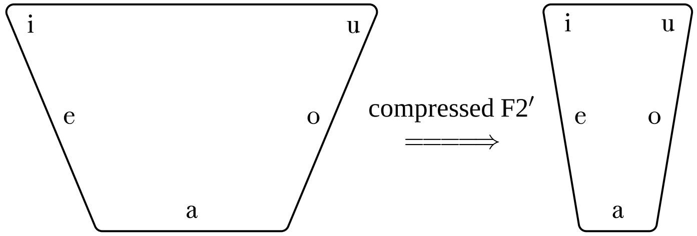
[*] Our deepest appreciation goes to audiences at MIT, the Stanford Phonology Workshop, the 44th Annual Meeting of the Chicago Linguistics Society, and the UCSC Linguistics Alumni Conference for their comments and suggestions, and to Adam Petrie for his invaluable help with various aspects of scripting in R.

[1] We follow Schwartz et al. in calculating Bark with the formula $f_{\mathrm{Bk}} = 7 \cdot \sinh^{-1}(f_{\mathrm{Hz}}/650)$.

Problem: Because F2′ spans a significantly larger range (about 10–11 Bk) than F1 does (only about 4–5 Bk), this simple euclidean measure of auditory distance overgenerates color (F2′) contrasts in comparison to height (F1) contrasts.

In order to generate more realistic predictions about color vs. height contrasts, phonetic models of vowel dispersion must compress the color space:

(4)



There is independent acoustic and perceptual support for weighting F1 more heavily than F2′. For example, F1 is known to be louder than higher formants, and louder formants weight more heavily in perceptibility (see Lindblom 1986, Schwartz et al. 1997a, Benkí 2003).

The amount of weighting F2′ receives is represented in DFT by the parameter λ, which falls between 0 (for which dispersion is determined solely by F1) and 1 (for which F1 and F2′ contribute equally to dispersion).

In DFT, the total dispersion energy E_D (5) of a vowel system with N vowels is the sum of the inverse squares of the λ-weighted distances d_{ij} between every pair of vowels V$_i$ and V$_j$ in the system:

(5) $$E_D = \sum_{\substack{i=1,\ldots,N-1 \\ j=i+1,\ldots,N}} \frac{1}{d_{ij}^2} \quad \text{where } d_{ij} = \sqrt{(\mathrm{F1}_i - \mathrm{F1}_j)^2 + \lambda^2(\mathrm{F2}'_i - \mathrm{F2}'_j)^2}$$

$$\boxed{\text{lower } E_D \Leftrightarrow \text{more perceptually peripheral vowel system}}$$

2.2 Focalization

DFT additionally assumes that some vowels, so-called “focal vowels”, are preferred in vowel systems due to their own inherent acoustic qualities, irrespective of the relational role they play in the system as a whole.

Specifically, a focal vowel in DFT has one or more pairs of adjacent formants that are close together, causing the formants to enhance each other, and making the vowel more perceptually robust overall (Schwartz and Escudier 1987, 1989; cf. Stevens 1972).

The focalization energy E_F of a vowel system is the sum of the focalization energies for each vowel in the system.

Each individual vowel’s focalization energy is the negative sum of the inverse squares of the differences between adjacent formants:

(6) $$E_F = \alpha \sum_{i=1}^{N} \left(\frac{-1}{(\mathrm{F1}_i - \mathrm{F2}_i)^2} + \frac{-1}{(\mathrm{F2}_i - \mathrm{F3}_i)^2} + \frac{-1}{(\mathrm{F3}_i - \mathrm{F4}_i)^2} \right)$$

$$\boxed{\text{lower } E_F \Leftrightarrow \text{more focal vowels}}$$

The most focal vowels in DFT by far are [i] and [y], with others ranked roughly as in (7):

(7) low E_F high E_F
(high magnitude negative) (low magnitude negative)
[i y] < [ɪ] < [e] < [ʏ] < [ɛ] < [æ a ɑ] < [u ʊ ø œ o ɔ ɒ] < [ə ʌ] < [ɨ ɯ ɤ]
most focal least focal

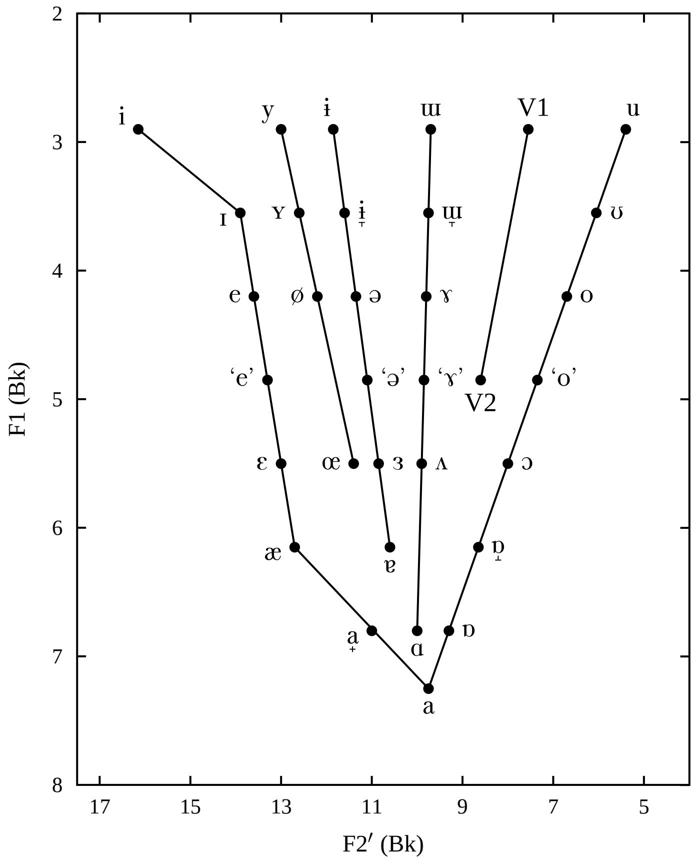
2.3 Prototypes

To limit the amount of computation time required to find optimal vowel systems, Schwartz et al. utilize a finite, predetermined set of 33 vowel “prototypes” (8).

These prototypes are based on the vowels appearing in UPSID, plus two extra abstract vowels [V1 V2] in the gap in the acoustic vowel space between the back round vowels and the back unrounded vowels. Though Schwartz et al. do not define these vowels physically, we take them to be back vowels with neutral lip positions (neither round nor unround, something like IPA [u̜ ‘o̜’] or [ɯ̹ ‘ɤ̹’]).

For each prototype vowel, Schwartz et al. set fixed values for F1–F4 that are typical of an adult male speaker, with F2′ calculated from F2, F3, and F4 by Mantakas et al.’s (1986) computation.

(8) *DFT prototypes*



2.4 Search algorithm

Searching the total set of possible systems to find the one single optimal system is not a computationally trivial task, even with the limitation of only having 33 vowel prototypes to choose from.

Thus, some search algorithm must be used which picks out only certain vowel systems for consideration of being the most optimal. We use the search algorithm in (9), an improvement over Schwartz et al.'s original search algorithm (see Sanders and Padgett 2008a for discussion):

(9) a. For each value of $N = 3, \ldots, 9$, initialize a catalog K_N of all vowel systems of size N already shown to be optimal by Schwartz et al. anywhere in the $\lambda \times \alpha$ space. For example:

$$K_5 = \left\{ \begin{array}{l} \text{[i e a ɔ u], [i y a ‘o’ u],} \\ \text{[i ‘e’ a ‘o’ u], [i ɛ a ɯ u]} \end{array} \right\}$$

b. For each value of N, randomly sample 5000 $\langle \lambda, \alpha \rangle$ pairs drawn from $[0,1] \times [0,1]$.[2]

c. For each $\langle \lambda, \alpha \rangle$ pair, randomly sample 4,603 candidate vowel systems of size N drawn from the 33 vowel prototypes (this is enough to have a 99% chance of finding a system in the top 0.1% of all possible systems in terms of optimality (lowest energy)). Add to this set of candidates all of the known optimal systems from K_N.

d. For each $\langle \lambda, \alpha \rangle$ pair and its set of candidate vowel systems, compute the energy of every candidate system, including those from K_N, according to equations (2,5,6).

e. For each $\langle \lambda, \alpha \rangle$ pair and its set of candidate vowel systems, select as optimal the candidate system with the lowest energy. If this optimal system is not yet in K_N, add it. Otherwise, make no change to K_N.

f. Repeat steps (b)–(e) five times, and then continue repeating them until K_N no longer changes.

This search algorithm explores the $\lambda \times \alpha$ space more thoroughly than Schwartz et al.'s original search algorithm, providing a more complete set of systems predicted to be optimal, for all N:

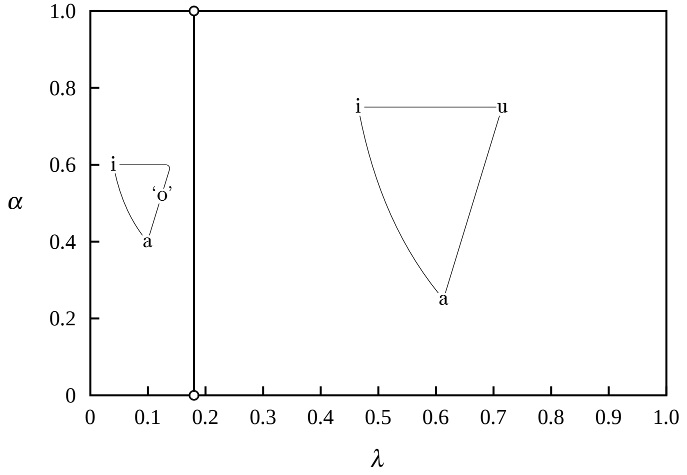
(10)

N	3	4	5	6	7	8	9
Schwartz et al. (1997a)	2	4	4	4	5	–	–
Sanders and Padgett (2008a)	7	10	11	11	9	13	10

To more easily visualize the optimal vowel systems that are found for various choices of $\langle \lambda, \alpha \rangle$, Schwartz et al. plot the optimal vowel systems in the $\lambda \times \alpha$ space by means of a “phase space” diagram, which divides the $\lambda \times \alpha$ space into regions where particular vowel systems are found to be optimal:

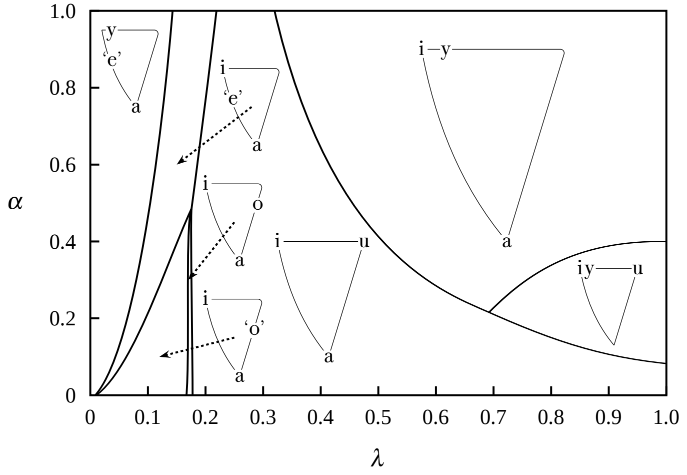
[2] Random sampling in this search algorithm was done using the `runif()` and `sample()` functions in the R programming language (Ihaka and Gentleman 1996).

(11) *Schwartz et al.'s phase space for* $N = 3$



The comparative phase space using Sanders and Padgett's search algorithm results in a more comprehensive map of the $\lambda \times \alpha$ space:

(12) *Sanders and Padgett's phase space for* $N = 3$



3 Articulation

3.1 Why articulation matters

Argument 1: The presence versus absence of a contrast affects "markedness". For vowel color, what counts as "unmarked" depends on how many contrastive vowel colors there are (Flemming 1995 [2002]):

3 vowel colors	2 vowel colors	1 vowel color
i ɨ u	i u	ɨ

Comparing three versus two colors, we might conclude that central vowels are more marked than front unround and back round vowels; i.e., *ɨ ≫ *i, *u.

However, if central vowels are truly the least marked vowel color, then it is odd that they show up precisely when a vowel system only has one vowel color, as in so-called "vertical" vowel systems like Kabardian (Choi 1989, 1991) and Marshallese (Choi 1995), where it seems *i, *u ≫ *ɨ.

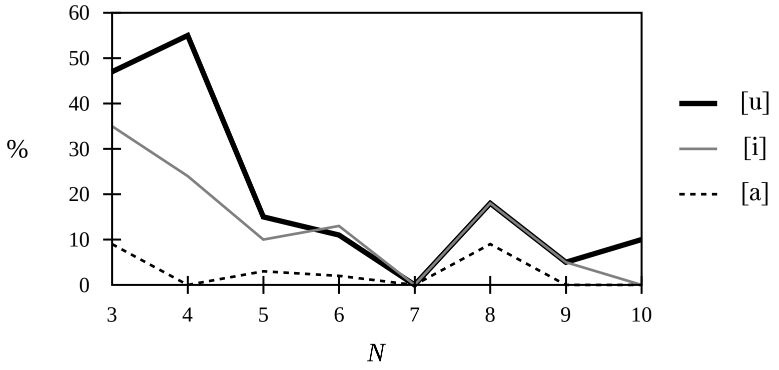
We find a similar asymmetry with the markedness of [ə], which is relatively marked when many contrasts exist (*ə ≫ *i, *u, *a), but in the absence of contrast (e.g., in reduction contexts), [ə] is common (*ə ≫ *i, *u, *a).

Argument 2: DFT's "transparency hypothesis". DFT generally does poorly at generating vowel systems with [ə]. For example, [i e ə a o u] is a relatively common type of 6-vowel system that DFT can't predict as optimal.

Hence Schwartz et al. (1997b) resort to a "transparency" rule for [ə], essentially stipulating that any DFT-generable system plus [ə] is a good system. Factoring articulatory ease into DFT's equations might render [ə] directly generable within DFT.

Argument 3: There seems to be a relationship between number of vowels and extremity of articulation. For example, our preliminary statistics on the absence of the "corner" vowels [i], [a], [u] in relationship to system size show a general downward trend (as well as an interesting asymmetry among the three vowels, with [u] missing more frequently and [a] missing less frequently):

(13) *Percentage of N-vowel systems in UPSID missing [u], [i], and [a]*



Conclusion: Languages avoid articulatory extremes when they are not necessary, and this should be encoded directly into DFT as a third energy component.

3.2 Adding articulation to DFT: First attempt

We propose a simple modification to the basic DFT energy equation, adding in a term for articulatory energy E_A:

(14) $E_{DFA} = E_D + E_F + E_A$

where E_A is given by the sum of the individual articulator energies of each vowel in the system (L for lips, H for tongue height, and B for tongue backness):

(15) $E_A = \gamma \sum_{i=1}^{N} (L_i + H_i + B_i)$

$\boxed{\text{lower } E_A \Leftrightarrow \text{more mid/central/neutral vowels}}$

After some initial testing, we started full simulations with a setting of $\gamma = 0.1$ and with the individual vowels' articulatory energies in (16):

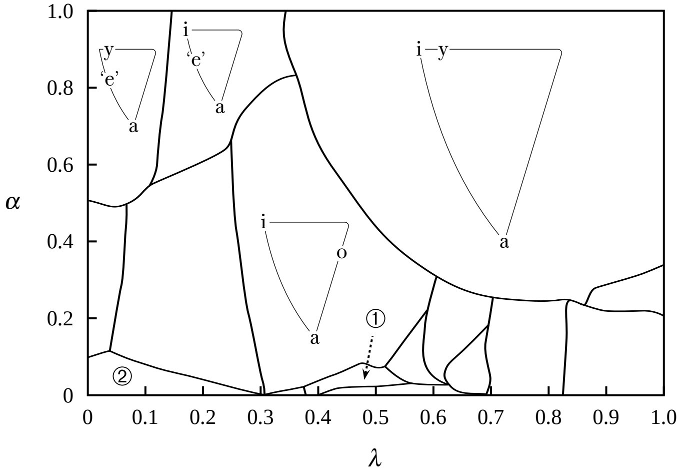
(16)	$L = 0$	neutral vowels	[ɨ ɨ̞ ə ‘ə’ ɜ ɐ a V1 V2]
	$L = 1/2$	round and spread vowels	[i y ɯ u ɪ ʏ ɯ̞ ʊ e ø. . .]
	$H = 0$	mid vowels	[‘e’ ‘ø’ ‘ə’ ‘ɤ’ V2 ‘o’]
	$H = 1/3$	upper-mid and lower-mid vowels	[e ø ə ɤ o ɛ œ ɜ ʌ ɔ]
	$H = 2/3$	near-high and near-low vowels	[ɪ ʏ ɨ̞ ɯ̞ ʊ æ ɐ ɒ̝]
	$H = 1$	high and low vowels	[i y ɨ ɯ V1 u a̟ a ɑ ɒ]
	$B = 0$	central vowels	[ɨ ɨ̞ ə ‘ə’ ɜ ɐ a]
	$B = 1/2$	front and back vowels	[i y ɯ V1 u ɪ ʏ ɯ̞ ʊ. . .]

So, the E_A for [i] would be $\gamma(1/2 + 1 + 1/2) = 0.2$, while [‘ə’] has $E_A = 0$.

3.3 Comparison of results

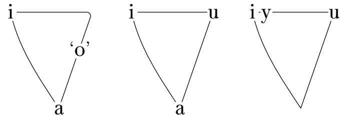
For $N = 3$, adding this articulatory parameter preserves four of the seven original systems found to be optimal in DFT (Sanders and Padgett 2008b). These four are shown in full in (17). Of these four systems, none are attested directly in UPSID, though [i a o] is similar to the attested vowel system [i a ‘o’] found in Pirahã. (The other three seem unlikely to represent real vowel systems.)

(17) *DFT+artic phase space for $N = 3$*



Note however that DFT originally *did* predict [i a ‘o’], but this system, along with [i a u] and [i y u], is lost when our articulatory parameter is added:

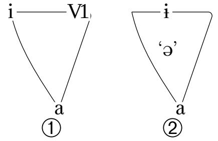
(18) *Missing predictions for* $N = 3$



Even though the loss of [i a ‘o’] is somewhat troubling, [i a o] is still retained, and the basic system type is still predicted to be optimal.

More problematic is the loss of [i u a], the most common 3-vowel system in UPSID, attested in at least 11 languages, including Tsimshian, Aleut, and Arrernte. The closest system predicted by DFT+artic is the system ①[i a V1], one of two new interesting systems in the revised model:

(19) *New interesting predictions for* $N = 3$



Many of the attested [i a u] systems may in fact have a more neutral vowel like [V1] than a fully rounded [u], but it seems unlikely that every single one of them has [V1] instead of [u], so the model definitely needs to be modified.

However, something is being done correctly: the newly predicted vowel system ①[i a V1] is a step closer to [i a ɯ], the vowel system of Jaqaru, which DFT does not originally predict.

Furthermore, the newly predicted vertical system ②[ɨ ‘ə’ a] is very promising. Nothing like it is predicted by DFT originally, and though this is not attested in UPSID, it is a vertical system of the kind mentioned in §3.1, attested in Kabardian and Marshallese.

The results are similarly mixed for larger vowel systems, with no apparent quantitative gain in number of system types predicted. In (24), “hits” are exact matches between prediction and attestation; “near hits” match all but one vowel, which is off by one position in the acoustic space; and “near-ish” hits match all but two vowels, which are each off by one position:

(20)

$N = 3$	hits	near hits	near-ish	total
DFT	2	1	1	4
DFT+artic	1[3]	4	1	6
$N = 4$				
DFT	4	6	1	11
DFT+artic	3	4	3	10
$N = 5$				
DFT	3	6	7	16
DFT+artic	2	5	10	17
$N = 6$				
DFT	1	2	6	9
DFT+artic	1	2	2	5
$N = 7$				
DFT	0	1	4	5
DFT+artic	0	0	7	7
$N = 3–7$				
DFT	10	16	19	45
DFT+artic	7	15	23	45

Qualitatively, however, there is a noticeable difference. Most of DFT+artic’s lost systems are similar to systems it does predict, but it also predicts new systems that belong to entirely different classes than what could be achieved in ordinary DFT, e.g., vertical 3-vowel systems and 5-vowel systems containing central vowels.

3.4 Adding articulation to DFT: Second attempt

From pilot simulations with different values of γ (the relative weight of the articulatory energy) and different maximum values of L and B (the individual

[3] Includes Kabardian and Marshallese, which are not in UPSID.

energies due to lip rounding and tongue backness), we found that the results were improved when γ was lower for higher values of N. Thus, we propose that instead of having E_A be the *sum* of the articulatory energies in a vowel system, it should be the *average*. This seems intuitively correct: a Spanish speaker isn't necessarily using more articulatory effort to speak than an Arabic speaker, simply because Spanish has more vowels.[4]

(21) $$E_A = \frac{\gamma}{N}\sum_{i=1}^{N}(L_i+H_i+B_i)$$

$$\boxed{\text{lower } E_A \Leftrightarrow \text{more mid/central/neutral vowels}}$$

Furthermore, we found that we still weren't getting quite as many systems with central vowels as we would expect based on UPSID, so we increased the maximum cost of L and B (though they were still kept less than the maximum value of H, which was held stable at 1.0).

We ran new full simulations with two settings for γ (0.27 and 0.24) and three settings for the maximum values of $L = B$ (0.8, 0.7, and 0.6).

3.5 Comparison of new results

For some combinations of values of γ and $L = B$, the new results using average articulatory energy show improvement over the first attempt using the sum of articulatory energy, and significant improvement over plain DFT:

(22) *Comparison of results for* $N = 3$ *across three DFT models*

model	γ	$L = B$	hits	near hits	near-ish	total
plain DFT	–	–	2	1	1	4
$+E_A$ sum	0.10	0.5	1	4	1	6
$+E_A$ avg	0.27	0.8	1	0	0	1
	0.27	0.7	2	4	1	7
	0.27	0.6	2	4	1	7
	0.24	0.8	3	2	1	6
	0.24	0.7	3	3	1	7
	0.24	0.6	3	3	1	7

The most promising settings here occur when $\gamma = 0.24$, not only because of the higher number of matching systems with UPSID, but also because the desirable system [i a u] is directly predicted (it is only a near-hit at best when $\gamma = 0.27$).

The results for $N = 4$ point us in a different direction, preferring $\gamma = 0.27$:

(23) *Comparison of results for* $N = 4$ *across three DFT models*

model	γ	$L = B$	hits	near hits	near-ish	total
plain DFT	–	–	4	6	1	11
$+E_A$ sum	0.10	0.5	3	4	3	10
$+E_A$ avg	0.27	0.8	3	4	3	10
	0.27	0.7	4	4	4	12
	0.27	0.6	4	4	4	12
	0.24	0.8	3	4	3	10
	0.24	0.7	3	4	3	10
	0.24	0.6	4	4	3	11

The main qualitative results are that the model with average articulatory energy slightly improves upon predictions of systems with mid-central vowels, but is slightly worse at predicting systems with high central vowels.

For $N = 5$, we find mild improvement almost across the board:

(24) *Comparison of results for* $N = 5$ *across three DFT models*

model	γ	$L = B$	hits	near hits	near-ish	total
plain DFT	–	–	3	6	7	16
$+E_A$ sum	0.10	0.5	2	5	10	17
$+E_A$ avg	0.27	0.8	3	5	10	18
	0.27	0.7	3	5	8	16
	0.27	0.6	3	7	8	18
	0.24	0.8	3	7	8	18
	0.24	0.7	3	7	7	17
	0.24	0.6	3	5	9	17

The main qualitative result is that the model with average articulatory energy improves upon predictions of systems with two low vowels, due to the increased ability to predict vowels like [ɐ] (which appears alongside [ɑ] in Koya).

[4] In fact, the same argument could apply to focalization. We plan to explore this idea in future work.

4 Wrap-up and future work

Adding an articulatory parameter to DFT, especially an average over articulatory energy, yields improvement in both quantitative and qualitative predictions, based on matches with attested vowel systems in UPSID, at least for $N = 3$–5.

Most notably, we get more predicted systems containing central vowels, especially [ə]. This includes vertical vowel systems.

In future work, we would like to test different settings for γ, L, and B (and perhaps H), catalog predictions for $N > 5$, and explore better metrics of “nearness” for counting hits.

References

Benkí, José R. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60:129–157.

Carlson, Rolf, Gunnar Fant, and Björn Granström. 1975. Two-formant models, pitch and vowel perception. In Gunnar Fant and Mark A. A. Tatham, eds. *Auditory Analysis and Perception of Speech*. London: Academic Press. 55–82.

Carlson, Rolf, Björn Granström, and Gunnar Fant. 1970. Some studies concerning perception of isolated vowels. *STL-QPSR* 11:19–35.

Choi, John-Dongwook. 1989. Phonetic evidence for a three-vowel system in Kabardian. *The Journal of the Acoustical Society of America* 86:S18.

——. 1991. An acoustic study of Kabardian vowels. *Journal of the International Phonetic Association* 21:4–12.

——. 1995. An acoustic-phonetic underspecification account of Marshallese vowel allophony. *Journal of Phonetics* 23.

de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28:441–465.

Flemming, Edward. 1995 [2002]. *Auditory Representations in Phonology*. Doctoral dissertation. University of California, Los Angeles. [Published in 2002. New York: Routledge].

Ihaka, Ross and Robert Gentleman. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5:299–314.

Liljencrants, L. and Björn Lindblom. 1972. Numerical simulations of vowel quality systems: The role of perceptual contrast. *Language* 48:839–862.

Lindblom, Björn. 1986. Phonetic universals in vowel systems. In John J. Ohala and Jeri J. Jaeger, eds. *Experimental Phonology*. Orlando: Academic Press. 13–44.

Maddieson, Ian. 1984. *Patterns of Sounds*. Cambridge: Cambridge University Press.

Maddieson, Ian and Karen Precoda. 1989. Updating UPSID. *UCLA Working Papers in Phonetics* 74:104–111.

Mantakas, M., Jean-Luc Schwartz, and P. Escudier. 1986. Modèle de prédication du ‘deuxième formant effectif’ F'_2—application à l’étude de la labialité des voyelles avant du français. In *Proceedings of the 15th journées d’étude sur la parole*. Société Française d’Acoustique. 157–161.

Sanders, Nathan and Jaye Padgett. 2008a. Predicting vowel inventories from a dispersion-focalization model: New results. In *Papers from the 44th Annual Meeting of the Chicago Linguistics Society*.

——. 2008b. Articulatory parameters in a dispersion-focalization model of vowel systems. Talk given at UC Santa Cruz Linguistics Alumni Conference.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry. 1997a. The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25:255–286.

——. 1997b. Major trends in vowel system inventories. *Journal of Phonetics* 25:233–253.

Schwartz, Jean-Luc and Pierre Escudier. 1987. Does the human auditory system include a large scale spectral integration? In M. E. H. Schouton, ed. *The Psychophysics of Speech Perception*. Dordrecht: Martinus Nijhoff Publishers. 284–292.

——. 1989. A strong evidence for the existence of a large-scale integrated spectral representation in vowel perception. *Speech Communication* 8:235–259.

Stevens, Kenneth N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In E. E. Davis Jr. and P. B. Denes, eds. *Human Communication: A Unified View*. New York: McGraw-Hill. 51–66.