26 April 2008
The 44th Annual Meeting of the Chicago Linguistics Society

Predicting Vowel Inventories from a Dispersion-Focalization Model: New Results[*]

Nathan Sanders
Williams College
nsanders@williams.edu

Jaye Padgett
Univ. of California, 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 (e.g., 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* parameters:

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

b. **focalization**: maximization of the importance of "focal" vowels such as [i] and [y] (cf. the quantal vowels of Stevens 1972).

We report results of new vowel system simulations following the original DFT formulas of Schwartz et al.

However, our means of selecting candidate systems for comparison explores the search space more effectively, allowing for more thorough and accurate computation of DFT's predictions.

In particular, we find a greater number of optimal systems than originally reported for DFT by Schwartz et al., throughout the entire range of possible parameter settings in DFT.

This seems to be a good result, since many of the newly discovered optimal systems have analogues in the UPSID database (Maddieson 1984, Maddieson and Precoda 1989). However, in all cases the number of languages is small.

[*]Our deepest appreciation goes to the audience at the Stanford Phonology Workshop for their comments and to Adam Petrie for his invaluable help with various aspects of scripting in R.

2 How Dispersion-Focalization Theory works

In DFT, vowel systems are compared according to their total "energy". The lower a vowel system's energy is, the more optimal it is.

For a given vowel system $\{V_1, \ldots, V_N\}$, each vowel V_i is characterized by its first four formants $\langle F1_i, F2_i, F3_i, F4_i \rangle$, measured in Bark.[1]

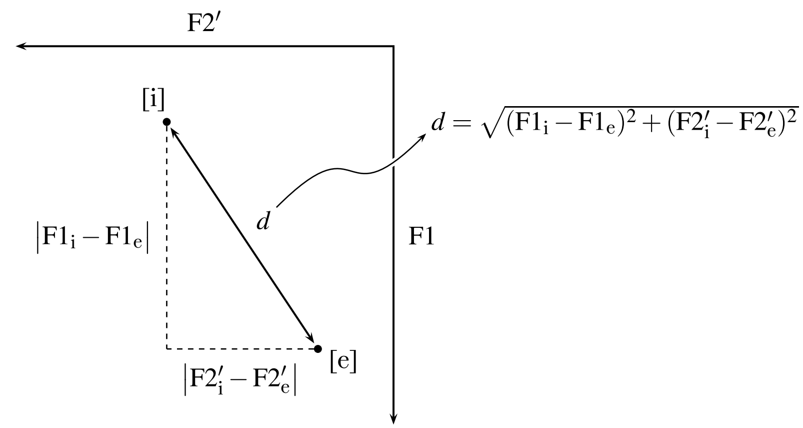
The system's total energy E_{DF} is the sum of its dispersion energy E_D and focalization energy E_F:

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

2.1 Dispersion

In DFT, the auditory distance d between two vowels V_i and V_j is the Euclidean distance between them in the auditory space based on their values of F1 and *effective* F2 (a.k.a. "F2 prime", a perceptual integration of F2, F3, and F4, symbolized as F2′), as demonstrated below:

(3)

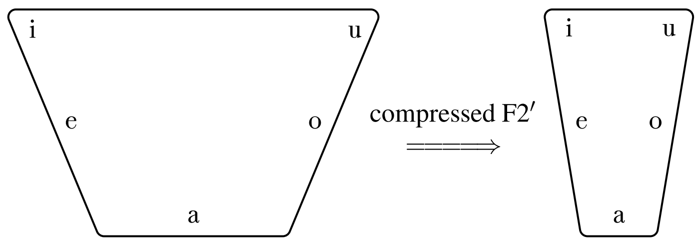


Problem: Because F2′ spans a significantly larger range (about 10–11 Bk) than F1 does (about 4–5 Bk), this simple measure for d overgenerates color (F2′) contrasts in comparison to height (F1) contrasts.

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

In order to generate more realistic predictions, 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′; e.g., F1 is louder than F2 (see Lindblom 1986, Schwartz et al. 1997a, Benkí 2003).

The amount of weighting F2′ receives is represented by the parameter λ, which falls between 0 (dispersion determined solely by F1) and 1 (equal weighting between F1 and F2′).

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

lower E_D ⇔ more peripheral vowel system

2.2 Focalization

DFT additionally assumes that some vowels, "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 (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) $$E_F = \alpha \sum_{i=1}^{N} \left(\frac{-1}{(\mathrm{F1}_i - \mathrm{F2}_i)^2} + \frac{-1}{(\mathrm{F2}_i - \mathrm{F3}_i)^2} + \frac{-1}{(\mathrm{F3}_i - \mathrm{F4}_i)^2} \right)$$

lower E_F ⇔ more focal vowels

The value of the parameter α determines the relative influence of *focalization* vis-à-vis *dispersion*.

The most focal vowels in DFT by far are [i] and [y]. Other vowels ranked by E_F include:

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

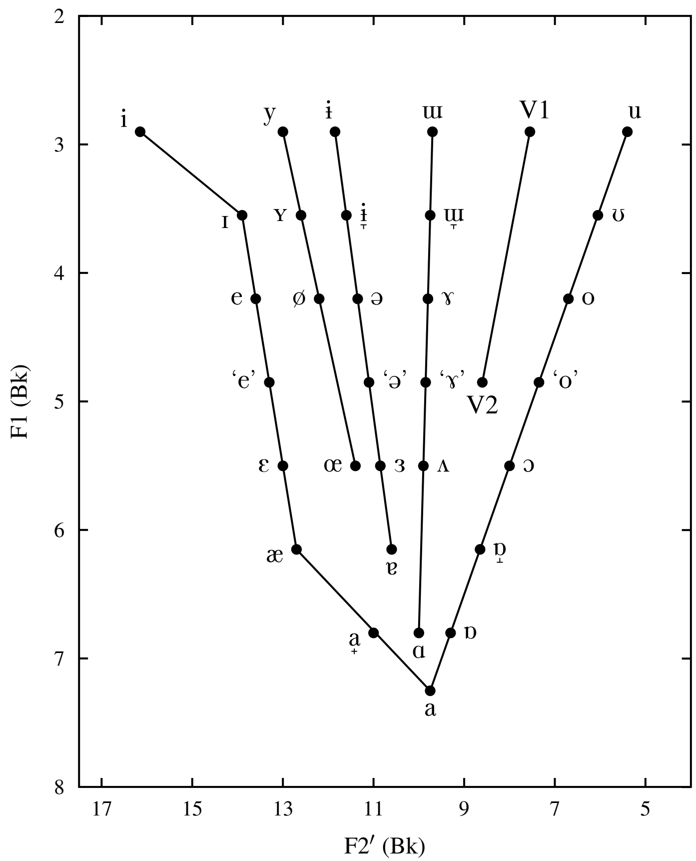
2.3 Prototypes

To limit the amount of computation required to find optimal vowel systems, Schwartz et al. utilize a finite, predetermined set of 37 vowel "prototypes" (8).

These prototypes are based primarily upon the vowel system from UPSID, plus a few extra vowels in the gap in the acoustic vowel space between the back round vowels and the back unrounded vowels.

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–F4 by Mantakas et al.'s (1986) computation.

(8) *DFT prototypes*



3 What Dispersion-Focalization Theory predicts

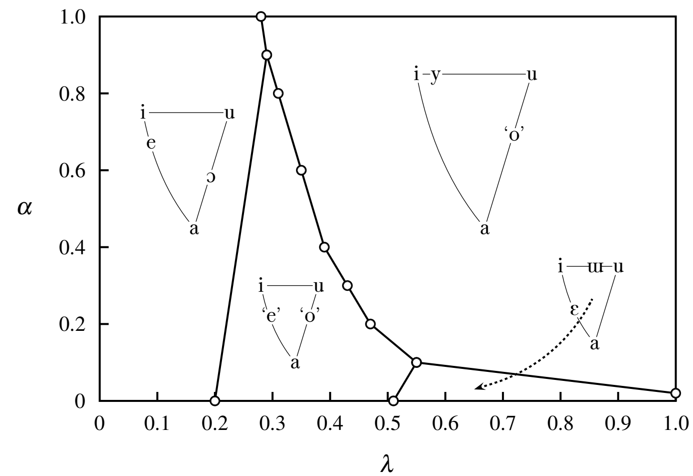
3.1 Phase spaces

The problem now is to find, for any number of vowels N, the optimal vowel system of size N, which minimizes the total energy E_{DF} (2).

The result will depend on the particular values of the two weighting parameters λ and α. Thus, there will be potentially different optimal vowel systems of size N for different choices of $\langle\lambda,\alpha\rangle$ pairs.

To visualize the results, Schwartz et al. plot the optimal vowel systems in the $\lambda\times\alpha$ space by means of a "phase space" plot, as in (9). This phase space shows that the vowel system [i e a ɔ u] is found to be optimal in the region to the far left, where $\lambda\leq 0.2$, as well as for some values $0.2\leq\lambda\leq 0.3$ as α increases.

(9) *Schwartz et al.'s phase space for* $N=5$



3.2 Search algorithm

Searching the total space of possible systems to find the one single optimal system is not a trivial task, even with the limitation of only having 37 vowel prototypes to choose from.

Thus, some search algorithm must be used which picks out only certain vowel systems to compare. Schwartz et al. utilize the following procedure:

(10) a. For each value of $N=3,\ldots,9$, pre-select a number of systems of size N that are candidates for being *the* optimal vowel system for some $\langle\lambda,\alpha\rangle$ in DFT.[2]

[2]Schwartz et al. do not specify precisely how these candidate systems are chosen, but presumably, they are hand-picked based on typological and phonetic plausibility. That is, [i e a o u] would most certainly have been a candidate 5-vowel system, but there would be no apparent need to even consider a system such as [ɨ ɯ ɤ ɘ ø], so it would likely have been completely ignored.

b. For various $\langle\lambda,\alpha\rangle$ pairs, compute the energy of every candidate system, according to equations (2,5,6).[3]

c. For each $\langle\lambda,\alpha\rangle$ pair, select as optimal the candidate system with the lowest energy.

Problem: There are at least two ways this search algorithm can go wrong:

(11) a. The initial posited candidate systems might not include the system that is actually the most optimal. Some truly optimal systems may be overlooked because they incorrectly appear to a human to be "obviously" less optimal than some other system, when in fact, the actual calculations prove otherwise.

b. The choice of $\langle\lambda,\alpha\rangle$ pairs may not be fine-grained enough. This would result in gaps in the phase space, which could cause some regions of the phase space to be presumed to be larger than they actually are and potentially miss out on finding some optimal systems. For example, had Schwartz et al. not considered any value of $\alpha < 0.1$, then the vowel system [i ɛ a ɯ u] would not have been found to be optimal, as can be seen in the phase space plot in (9).

3.3 Our revised search algorithm

Using the same prototypes and same functions as Schwartz et al., we used a different search algorithm for finding optimal systems and mapping out the phase space, in an effort to alleviate or avoid the problems in (11).

(12) 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, as per (9):

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

This initialized catalog only needs to be created once.

b. For each value of N, randomly sample 5000 $\langle\lambda,\alpha\rangle$ pairs drawn from $[0,1]\times[0,1]$.[4]

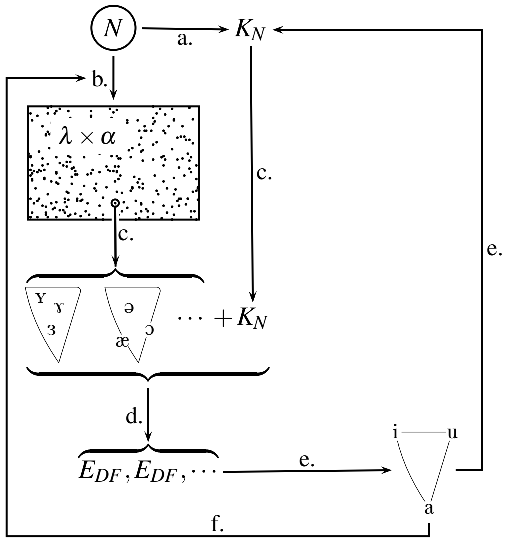
c. For each $\langle\lambda,\alpha\rangle$ pair, randomly sample enough candidate vowel systems of size N drawn from the 37 vowel prototypes to have a 99% chance of finding a system that is in the top 0.1% of all possible systems in terms of optimality (lowest energy).[5] Add to this set of candidates 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.

(13) *Schematic diagram for (12)*



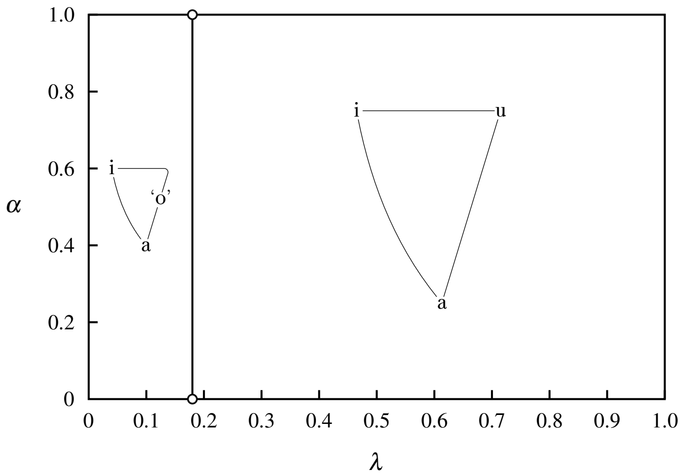
[3] Again, Schwartz et al. do not specify precisely how these $\langle\lambda,\alpha\rangle$ pairs are chosen. Given the locations of their key plotted points in their phase spaces, $\langle\lambda,\alpha\rangle$ seems to have been chosen largely at fixed intervals, though this appearance could just be an artifact of their graphing method.

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

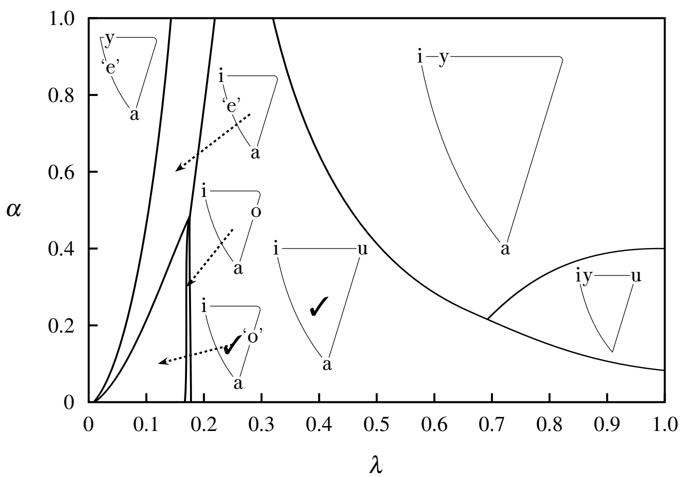
[5] See Appendix A for mathematical discussion of what constitutes "enough" candidates.

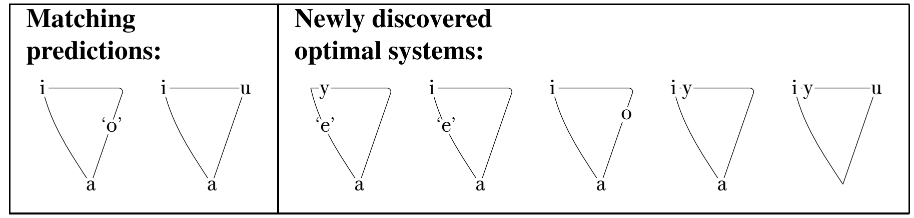
3.4 Comparison of results

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



(15) *New phase space for $N = 3$*





Of the five new optimal systems, [i a o] is optimal only in a tiny sliver of the phase space and is very similar to already predicted [i a 'o'], so this system is only marginally "new". The remaining four are not attested in UPSID and seem to be phonetically implausible.

Overall, while we do not find any definitively new optimal systems for $N = 3$ that are attested in UPSID, our results show that the parameters λ and α must be limited to the roughly triangular region that stretches from the entire bottom (where $\alpha = 0$) up to $0.2 \leq \lambda \leq 0.35$ for $\alpha = 1.0$.

There are still some 3-vowel systems attested in UPSID that do not yet appear to be optimal in DFT, even with our new search algorithm. For example:

[i æ u]	Shilha
[e a o]	Alabama and Amuesha
['ə' a 'o']	Qawasqar

Interlude: Evaluating the results using UPSID

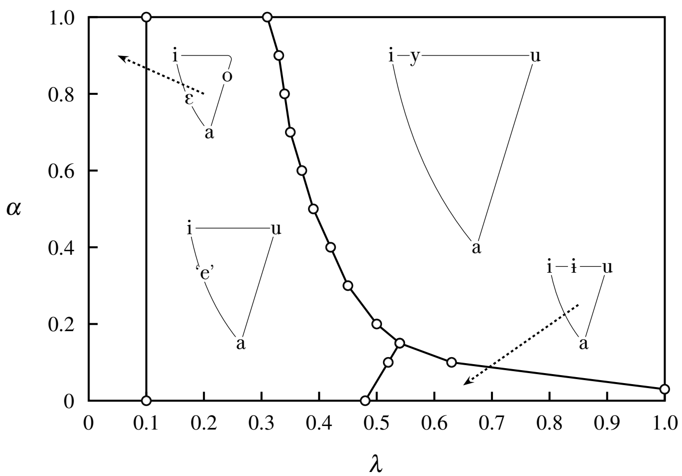
UPSID is an invaluable resource, but using it to evaluate predictions raises significant challenges.

Insufficient precision: When a researcher notates *e*, is this [e], ['e'], or [ɛ]? (UPSID often uses *'e'* when the source description doesn't make the precise height clear.) And so on for *u* ([u] or [ʊ]?), etc. Further, when a phoneme has allophones, which one is recorded in UPSID?

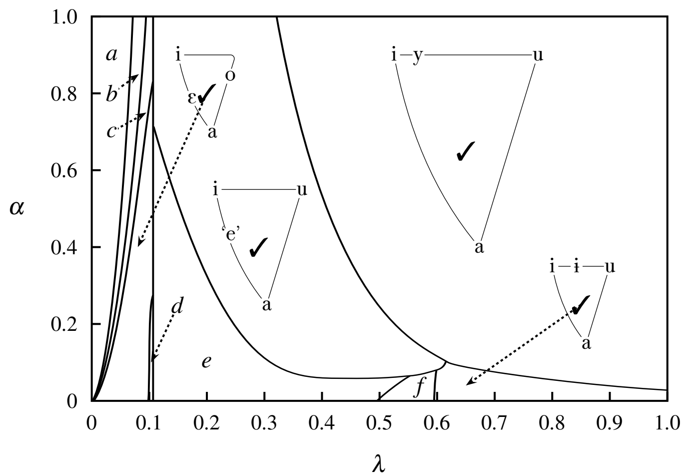
Too much precision: There are enough symbols used by UPSID that many vowel systems appear only once. This makes it hard to talk about "frequent" systems. Schwartz et al. (1997b) address this problem by "collapsing" various kinds of distinctions in UPSID and in DFT's outputs, using a representative vowel for a set of similar vowels when only one member is present:

[i ɪ] → [i]	[ɯ ɯ̞ u ʊ] → [u]	[ø 'ø' œ] → [ø]
[y ʏ] → [y]	[ɨ ɨ̞ ʉ ʉ̞] → [ɨ]	[ɤ 'ɤ' ʌ] → [ɤ]

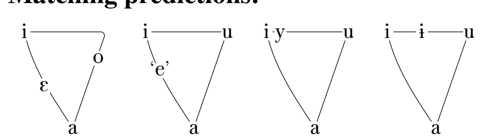
(16) *Schwartz et al.'s phase space for* $N = 4$



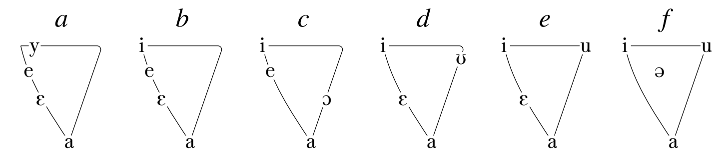
(17) *New phase space for* $N = 4$



Matching predictions:



Newly discovered optimal systems:



Two of these new optimal systems, ^{a}[y e ɛ a] and ^{b}[i e ɛ a], are unattested in UPSID and seem phonetically implausible.

The third new optimal system, ^{c}[i e a ɔ], is not attested directly in UPSID, but it is similar to attested systems, such as Klamath and Tacana's [i 'e' a 'o'].

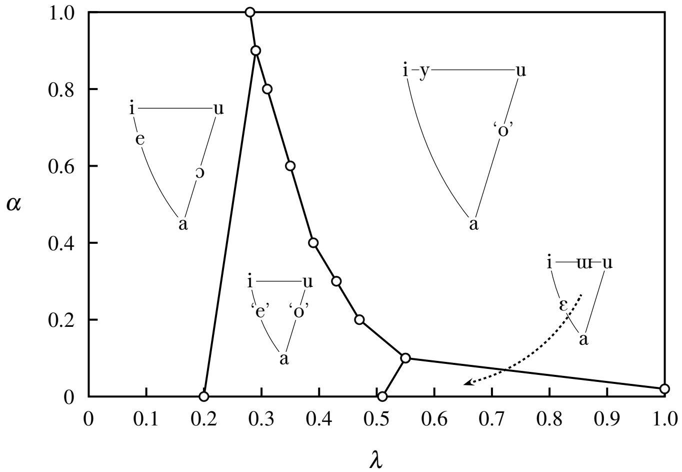
New matches with UPSID! The three remaining systems, ^{d}[i ɛ a ʊ], ^{e}[i ɛ a u], and ^{f}[i ə a u], are each attested in UPSID, either exactly (in Murinhapatha, Moxo, and Jebero, respectively) or very similarly (Cayapa's [i ɛ ɑ ʊ], Lushootseed's [ɪ 'ə' a ʊ], and Ivatan, Paiwan, and Yupik's [i 'ə' a u]). This shows that DFT's ability to predict extant systems is better than originally thought.

The system ^{f}[i ə a u] is also interesting because it contains [ə], which calls into question the need for Schwartz et al.'s (1997b) "transparency hypothesis" for [ə], in which neither the presence nor absence of [ə] plays a role in the acoustic dispersion of a vowel system. Boë et al. (1994) argue that [ə] is difficult to obtain in DFT, but as our results show, there is indeed a region of the phase space (albeit small!) around $\langle 0.55, 0.01\rangle$ where an optimal system containing [ə] is predicted.

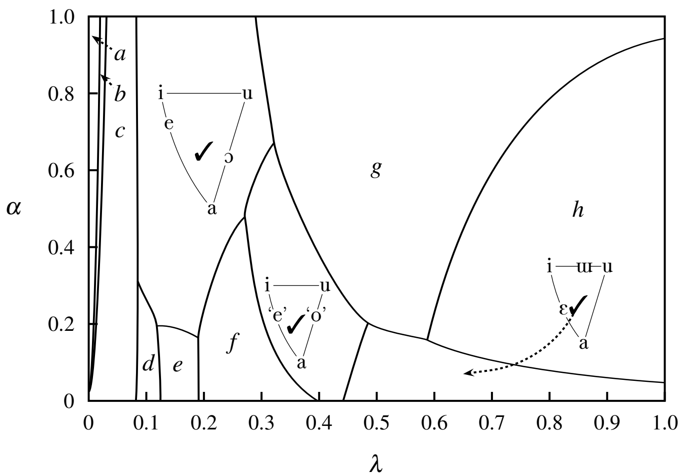
There are still numerous 4-vowel systems attested in UPSID that do not yet appear to be optimal in DFT. For example:

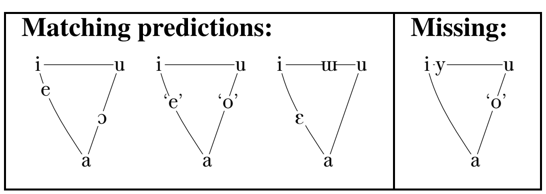
[i a ɔ u]	Tiwi	[ɨ ɛ a 'o']	Margi
[i æ a u]	Quileute	[ɨ ɜ a ɒ]	Yessan-Mayo
[ɪ ɐ a ɯ̞ɪ]	Nunggubuyu	['e' 'ə' a 'o']	Upper Chehalis
[i ɨ a 'o']	Maranao		

(18) *Schwartz et al.'s phase space for $N = 5$ (repeated from (9))*

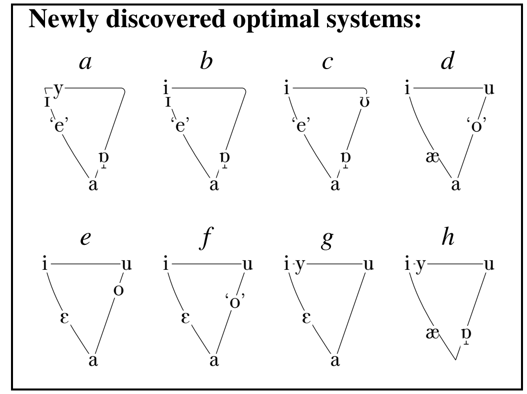


(19) *New phase space for $N = 5$*





The missing system [i y a 'o' u] doesn't seem to be a problem, because it is not attested in UPSID, and the relevant phase space region lies almost entirely in the upper right region ruled out by $N = 3$.



Of the eight new optimal systems, two systems, ^{a}[ɪ y 'e' a ɒ̞] and ^{b}[i ɪ 'e' a ɒ̞], are unattested in UPSID and seem phonetically implausible. Three other systems, ^{c}[i 'e' a ɒ̞ ʊ], ^{g}[i y ɛ a u], and ^{h}[i y æ ɒ̞ u], while potentially plausible, are also unattested in UPSID. All five of these systems are optimal only in phase space regions already ruled out from $N = 3$.

The system ^{d}[i æ a 'o' u] is not attested directly in UPSID, but it is similar to attested systems, such as Taishan's [i æ a ɔ u], and with a bit of a stretch, possibly Nez Perce's [ɪ æ a ɔ ɯ].

New matches with UPSID! The systems ^{e}[i ɛ a o u] and ^{f}[i ɛ a 'o' u] are both directly attested in UPSID (Jacaltec and Nasioi; and Batak, Hawaiian, Yucuna, Yagaria, and Baining).

There are still numerous 5-vowel systems attested in UPSID that do not yet appear to be optimal in DFT. For example:

[i ɨ a ɔ u]	Papago	[i e ɐ ɑ u]	Koya
[i ɨ 'e' a 'o']	Abipon	[i ø æ a ɤ]	Hopi
[i ɨ ɛ a 'o']	Cofan	[i ø̠ ɛ a ʊ]	Malakmalak
[i ɛ a ɔ o]	Tseshaht	['e' æ ɔ u ɯ]	Hixkaryana
[i e a ʊ u]	Kunimaipa		

Results for $N = 6, \ldots, 9$ are in Appendix B.

4 Wrap-up and future work

Our improved search algorithm leads to different results for DFT's predictions. In particular, we find a greater number of optimal systems throughout the entire $\lambda \times \alpha$ space for $N = 3, \ldots, 7$ (Schwartz et al. do not give details for $N = 8$ or 9):

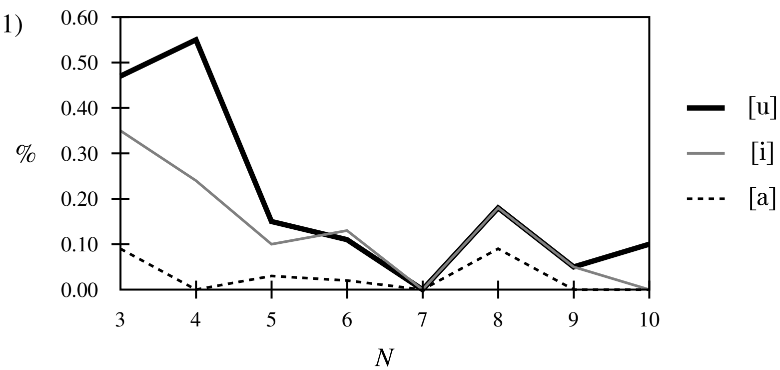
(20)

N	3	4	5	6	7	8	9
original	2	4	4	4	5	–	–
new	7	10	11	11	9	13	10

Many of these newly discovered optimal systems are attested in UPSID, which shows that the predictive power of DFT is greater than originally thought.

Preliminary statistics on the percentage of systems missing one of the "corner" vowels ([i], [u], and [a]) show an interesting asymmetry:

(21)



An articulatory explanation may help here, so one of our next steps is to add an articulatory energy component to the basic DFT equation, with more articulatorily extreme vowels having higher articulatory energy. This would also increase the relative optimality of systems containing vowels like [ə], further obviating Schwartz et al.'s (1997b) transparency rule for [ə].

Finally, further thought needs to be given to precisely what it is that DFT is trying to predict. Are we only concerned with truly optimal systems, or should relative optimality tie into relative frequency? Answering this question may be necessary to incorporate "crazy" vowel systems (such as Qawasqar's [ə a 'o']) into the DFT model.

Appendix A: How many candidates do we need?

In order to have a 99% chance of randomly selecting one of the systems in the top 0.1% of all possible systems in terms, we need to randomly sample 4,603 candidate systems.

Proof: Let S be the total number of systems to be randomly sampled. The probability that not a single one of them is in the top 0.1% is also the probability that all of them are in the bottom 99.9%. Each one has a probability of 0.999 of being in the bottom 99.9%, so the probablity that all of them are in the bottom is all S of those probabilities of 0.999 multiplied together: 0.999^S.

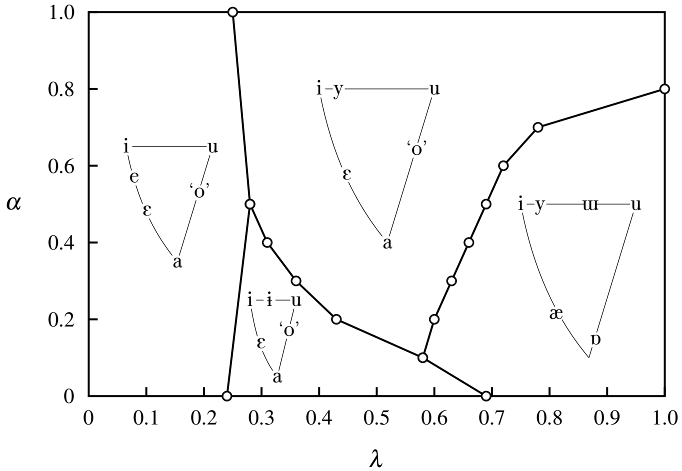
Having *all* of our sampled systems be in the bottom 99.9% would be a bad result, so we want the probability of this event occurring to be as small as possible. A 1% chance of failing to get any optimal system seems acceptable, so we set $0.999^S = 0.01$ and solve for S:

$$\begin{aligned} 0.999^S &= 0.01 \\ S \cdot \log(0.999) &= \log(0.01) \\ S &= \frac{\log(0.01)}{\log(0.999)} = 4{,}603 \end{aligned}$$

We actually perform better than a 1% failure rate, due to the addition of known optimal systems from the catalog to the 4,603 random candidates, due to the repetition in step (12f), and due to the genetic nature of the algorithm, which causes the catalog to grow every time a new optimal system is found.

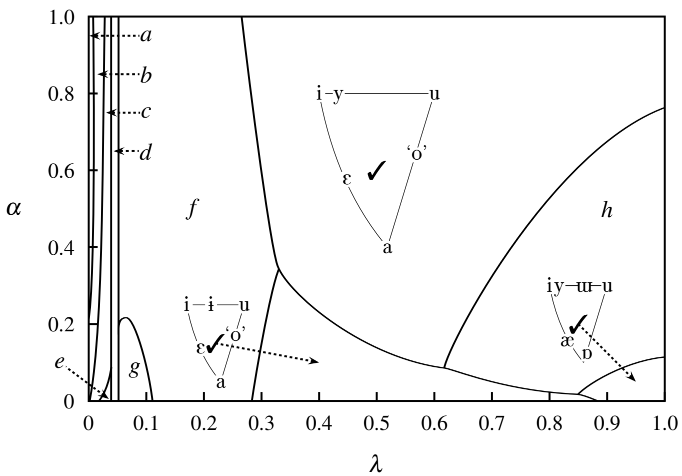
Appendix B: Predictions for *N* = 6,. . . ,9

(22) *Schwartz et al.'s phase space for* $N = 6$



[i e ɛ a 'o' u] [i ɨ ɛ a 'o' u] [i y ɛ a 'o' u] [i y æ ɒ ɯ u]

(23) *New phase space for* $N = 6$

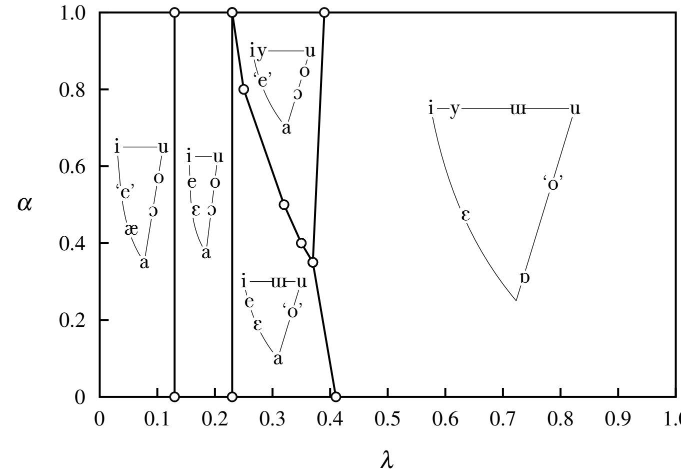


[a][ɪ y e ɛ æ a]

[b][i ɪ e ɛ a ɒ̞] [d][i e æ a 'o' ʊ] [f][i e æ a 'o' u] [i y ɛ a 'o' u] [h][i y æ a 'o' u]

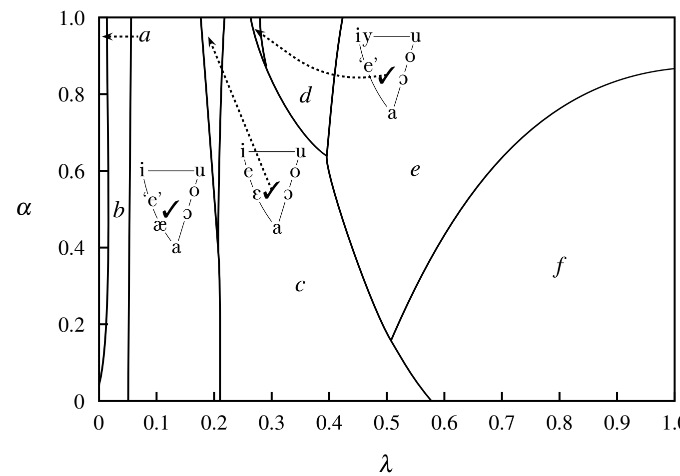
[c][i e a ɒ̞ 'o' ʊ] [e][i ø ɛ a ɒ̞ ʊ] [g][i 'e' a ɒ̞ o u] [i ɨ ɛ a 'o' u] [i y æ ɒ ɯ u]

(24) *Schwartz et al.'s phase space for* $N = 7$



[i 'e' æ a ɔ o u] [i e ɛ a ɔ o u] [i e ɛ a 'o' ɯ u] [i y 'e' a ɔ o u] [i y ɛ ɒ 'o' ɯ u]

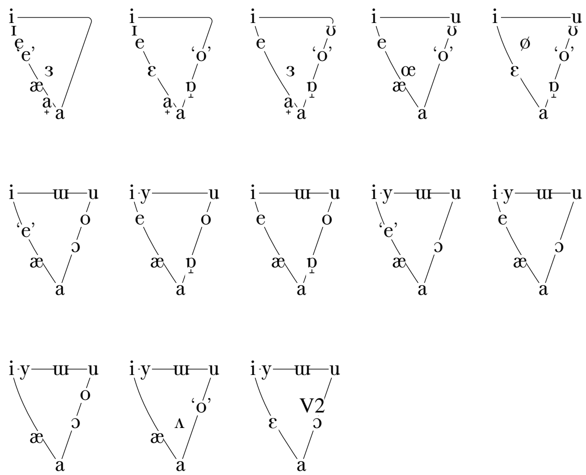
(25) *New phase space for* $N = 7$



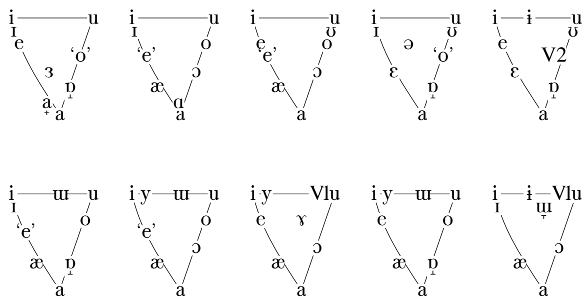
[a][i e ɛ a ɒ̞ 'o' ʊ] [i 'e' æ a ɔ o u] [i e ɛ a ɔ o u] [i y 'e' a ɔ o u] [e][i y ɛ a 'o' ɯ u]

[b][i ɪ e ɛ æ a V1] [c][i e æ a 'o' ɯ u] [d][i y 'e' æ a ɔ u] [f][i y æ a 'o' ɯ u]

(26) *Predicted optimal systems for* $N = 8$



(27) *Predicted optimal systems for* $N = 9$



References

Benkí, José R. 2003. Analysis of English nonsense syllable recognition in noise. *Phonetica* 60.2:129–157.

Boë, Louis-Jean, Jean-Luc Schwartz, and Nathalie Vallée. 1994. The prediction of vowel systems: Perceptual contrast and stability. In E. Keller, ed. *Fundamentals of Speech Systhesis and Speech Recognition*. Chinchester, UK: John Wiley. 185–213.

de Boer, Bart. 2000. Self-organization in vowel systems. *Journal of Phonetics* 28.4:441–465.

Ihaka, Ross and Robert Gentleman. 1996. R: A language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5.3:299–314.

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, MA: 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.

Schwartz, Jean-Luc, Louis-Jean Boë, Nathalie Vallée, and Christian Abry. 1997a. The Dispersion-Focalization Theory of vowel systems. *Journal of Phonetics* 25.3:255–286.

——. 1997b. Major trends in vowel system inventories. *Journal of Phonetics* 25.3:233–253.

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.