Explaining vowel inventory tendencies via simulation: finding a role for quantal locations and formant normalization

Brian Roark

Brown University

1. Introduction

This paper presents some new vowel inventory simulations, in the tradition of Liljencrants and Lindblom (1972) and Lindblom (1986) (henceforth LL72 and L86, respectively), papers which attempted to predict likely vowel configurations based on a maximization of perceptual contrast. The predictions made by our simulations better fit actual vowel inventory tendencies by building certain fundamental, heretofore ignored features of vowel perception and production into the vowel space model. In particular, we build quantal locations into the model, which are privileged in terms of both perceptual salience and resistance to articulatory imprecision (Stevens, 1972). Neglecting this factor caused the two earlier models to make erroneous predictions for inventories with as few as four vowels. In a second simulation, we incorporate formant normalization, which is a likely perceptual process given broad speaker formant range variation (see, e.g., Nearey, 1978). This results in a preference for peripheral vowels, even in higher inventories, unlike in previous simulations.

The notion of contrast, and its phonetic motivation, is becoming increasingly important in phonological theory as it is practiced in certain circles. Many recent papers (e.g., Flemming, 1996; Silverman, 1998; Hayes, 1999) have cited the vowel inventory simulations of LL72 and L86 as instances of the phonetic grounding of phonologically interesting generalizations: namely, why particular vowel inventory configurations arise more often than others. Insofar as simulations make accurate predictions about likely configurations of vowel inventories based on a well-motivated definition of perceptual contrast, they provide support for approaches that reserve a key role for contrast (e.g.

contrast maintenance) in phonological processes.

Despite the influential status of LL72 and L86, the predictions they make do not in every case correspond to the attested patterns of inventories in the world's languages (see, e.g., Schwartz et al., 1997). In particular, in inventories with four vowels, they failed to find the most common configuration; and as the size of the inventories grew, they predicted more non-peripheral vowels than are typically found in languages. These discrepancies can be summed up in two cross-linguistic tendencies which their simulations failed to incorporate: that the corners of the vowel space are almost universally occupied; and that peripheral vowels (i.e. front, back, and low) are far more common than non-peripheral vowels.

This paper is a re-examination of vowel inventory simulations in an attempt to make better predictions by bearing three questions in mind. First, how can we best represent the vowel space to capture the true nature of contrast within it? To this end, we examine the relative importance of the first and second formants as dimensions of contrast within the model, as well as the inclusion of certain uncontroversially fundamental features of vowel production and perception. Second, what counts as “distance” in the space that we define, and how does this relate to explicit variations in the representation of the space? Finally, how do we evaluate the results of the simulations, particularly given the nature of the search algorithm that is used, which is prey to local maxima?

Ultimately what we hope to learn from this exercise is whether some well-motivated definition of contrast will serve to explain why certain configurations arise more frequently in the world's languages than others, and what the dimensions of this contrast are. It could very well be that contrast is not simply distance in an acoustic or perceptual space, rather something tied up with the way in which formant frequencies and vowel patterns are recognized.

The rest of the paper will be organized as follows. First we will provide some background in the acoustics of vowels, and an introduction to the assumptions and practice underlying the simulations that we will be performing. Next we will review the attested cross-linguistic tendencies of vowel inventories. This will be followed by a detailed discussion of simulation considerations, during which we will motivate the decisions that were made for the simulations that will be reported here. Finally, we will present the simulations and results, and discuss the degree to which they improve upon the LL72 and L86 simulations.

2. Background

Several dimensions of contrast can be used to distinguish vowels acoustically. Of primary importance are the formant frequencies, in particular the first three. Formant frequencies are those frequencies at which there is a spectral peak, i.e. the frequencies at which the configuration of the vocal tract allows a maximum amount of energy to emerge, relative to neighbor frequencies. The first formant is the formant with the lowest frequency, the second formant at the next lowest frequency, and so on. As will be seen when we discuss the cross-linguistic tendencies below, once the vowel inventory reaches a certain size (around 9), additional dimensions of contrast are typically required, e.g. length distinctions or nasality. For the purpose of this study, we will be concentrating on inventories of size less than or equal to nine vowels, so these additional distinctions will be ignored. Also, for simplicity, we will be considering just the first and second formants[1]. The third formant does provide information, yet its range of variation, and hence its influence on contrast, is limited compared to the first two formants. Our approach would be easily extendible to include third formants, and this may prove an interesting extension, but for now we will be considering just the first and second formants as the dimensions of contrast.

Quantal vowels (Stevens, 1972) are those where two of the spectral peaks (formant frequencies) are very close to one another. This is preferred in two distinct

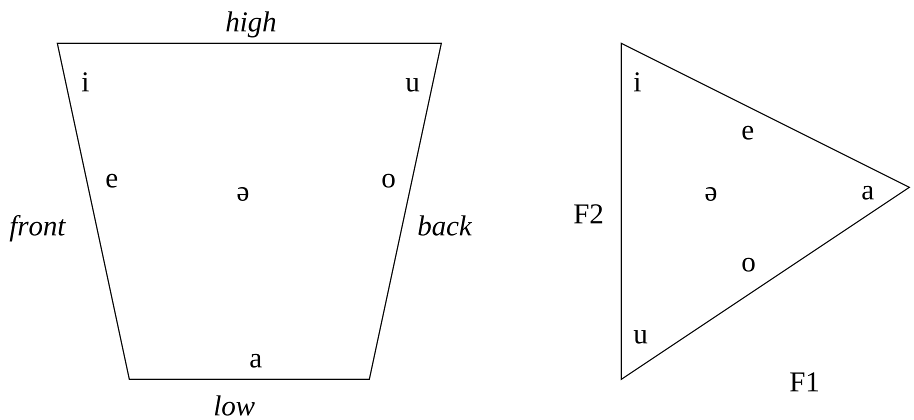

Figure 1. Vowel space representations: standard phonological grid and formant space

ways. First, such a configuration is perceptually salient, insofar as there is a broader peak, consisting of the two peaks in aggregate. Second, these vowels allow for more articulatory imprecision, since the formants change very slowly in their neighborhood of articulation. For these reasons, quantal vowels should be seen as privileged locations in the vowel space.

We have been speaking of the formant space as though it is fixed and universal, when in fact it varies from speaker to speaker. Formant frequencies are a property of the configuration of the vocal tract: large vocal tracts will have lower formant frequencies than small vocal tracts, i.e. the range of formant frequencies for an adult male is typically quite different from that of a child. Vowel perception involves some kind of normalization (see e.g. Nearey, 1978), so that a child's /a/ can be perceived as the same vowel as an adult male's /a/, despite the fact that the actual formant frequencies differ. We can think of the vowel space representations that will be presented in what follows as the representation

[1] Quantal locations involving the convergence of higher formants will be included, hence there will some influence of higher formants on the model.

of a single person's space, which may vary with respect to where precisely it falls in frequency, but which retains the same basic spatial relationships.

The simulations that we will perform are quite similar to those performed in the previously cited studies. The underlying assumptions of such an enterprise can be summarized as follows: (i) there is a range of possible vowel locations that makes up a perceptual "space"; (ii) there is a tendency to maximize contrast between vowels within a particular inventory; (iii) contrast = distance in the perceptual space; and thus (iv) we can simulate contrast maximization by maximizing distance within the stipulated vowel space. This very general characterization covers the approaches in LL72 and L86, as well as our own approach. A simulation involves: (i) stipulating a vowel space; (ii) placing a specified number of vowels (an "inventory") randomly within the space; and (iii) iteratively moving the vowels in a direction that increases the total distance between the vowels, until no move increases the distance. The configurations that result from this method can then be compared to attested configurations in the world's languages. The tendencies of these attested configurations will be discussed next.

3. Cross-linguistic tendencies

Schwartz et al. (1997) contains a comprehensive survey of the vowel inventories of 317 languages in the UCLA Phonological Segment Inventory Database (UPSID), and this survey will serve as the empirical point of comparison for evaluating the results of our simulation[2]. Their analysis begins by dividing vowel inventories into primary and secondary systems, where primary systems contrast vowels via formant patterns, while secondary systems utilize some additional dimension for contrast, usually length/quantity or nasality. Here we will be considering only the primary systems, since we are looking at formant based contrast.

The generalizations that we will be discussing are given in terms of regions of the vowel space. Figure 1 gives two representations of the vowel space, that which is traditionally used in phonology, and an idealized formant space. The traditional space is labelled with the terms which describe the boundaries of the space. These terms are equally applicable to the formant space, as one can see from the relative locations of the displayed vowels in both spaces. Vowels are called peripheral if they fall on the front, back, or low boundary of the space, and are otherwise called non-peripheral.

There are three generalizations of primary systems mentioned in Schwartz et al. that we will attempt to account for: (i) vowels tend to be concentrated at the periphery of the system; (ii) there tends to be an equal number of front and back peripheral vowels; and (iii) if there is an asymmetry on the periphery, front vowels occur more frequently than back vowels. An additional generalization which we will try to account for, and which is present in the Schwartz et al. (1997) data, but which is not a new generalization, is that the three corners of the vowel space (i.e. the quantal vowels: /a/, /i/, and /u/), almost

[2] We will also be citing some data from Crothers (1978), which was the empirical touchstone for the L86 simulations.

always occur in any vowel inventory.

In order to evaluate the simulation results with respect to the attested configurations gathered from the UPSID database, we will take the configurations returned by the simulation and score them with respect to: (i) the number of low peripheral vowels; (ii) the number of front peripheral vowels; (iii) the number of back peripheral vowels; and (iv) the number of non-peripheral vowels. High corners of the vowel space will be counted as being on their respective periphery, front or back.

4. Simulating vowel configurations

As with any simulation, ‘good' performance is highly parameter dependent. In this section we will attempt to motivate a set of simulation parameters that are both simpler and, in some respects, richer than those used in the previous simulations. In order to understand how these parameters can influence the performance of the simulations, we must first outline the general simulation procedure.

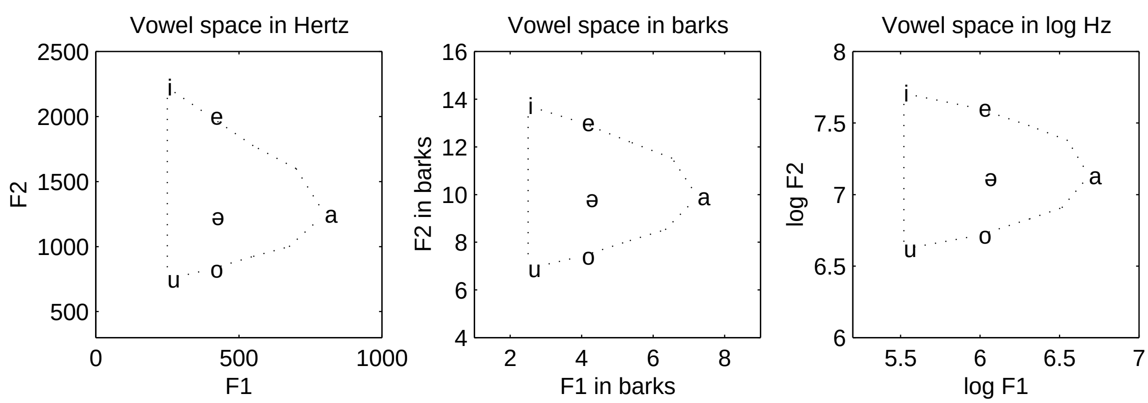

Figure 2. Representations of the vowel formant space in Hertz, Barks and Log Hertz

The general simulation procedure that will be used is the same as that used in the previous work. The vowel space is delimited, based upon what is articulatorily possible. Within the specified vowel space, a number of vowels are randomly placed. An algorithm then moves the vowel locations in such a way that it converges on a configuration where the distance between vowels in the space is locally maximized.

This formulation, however, leaves several specifics open to interpretation; most importantly: what is the vowel space, and what is distance within the vowel space. The vowel space was one difference between the LL72 and L86 simulations, the former being in Mels and the latter in Barks, which are each logarithmic transformations of the formant space that capture certain features of the auditory system. We will demonstrate that, while these units of measure are perhaps well motivated from a strictly perceptual standpoint, their use in simulations contained a bias towards F2 contrast over F1 contrast. If there is to be any bias, it should probably run in the other direction, since lower formants are typically more salient than higher formants. Our perspective, however, is that F1 and F2 are two independent dimensions of contrast (within the boundaries of the space), and should be considered equal in terms of their contribution to contrast in the vowel system.

Figure 2 shows the potential first and second formant vowel space in three different ways[3]. The first is in standard Hertz; the second in Barks; and the third in log Hertz. The thing to notice from these graphs is that the basic shape and configuration is quite similar in all cases, which may lead one to believe that it does not make much of a difference which unit of measure is used, particularly if we are primarily interested in regions and not points.

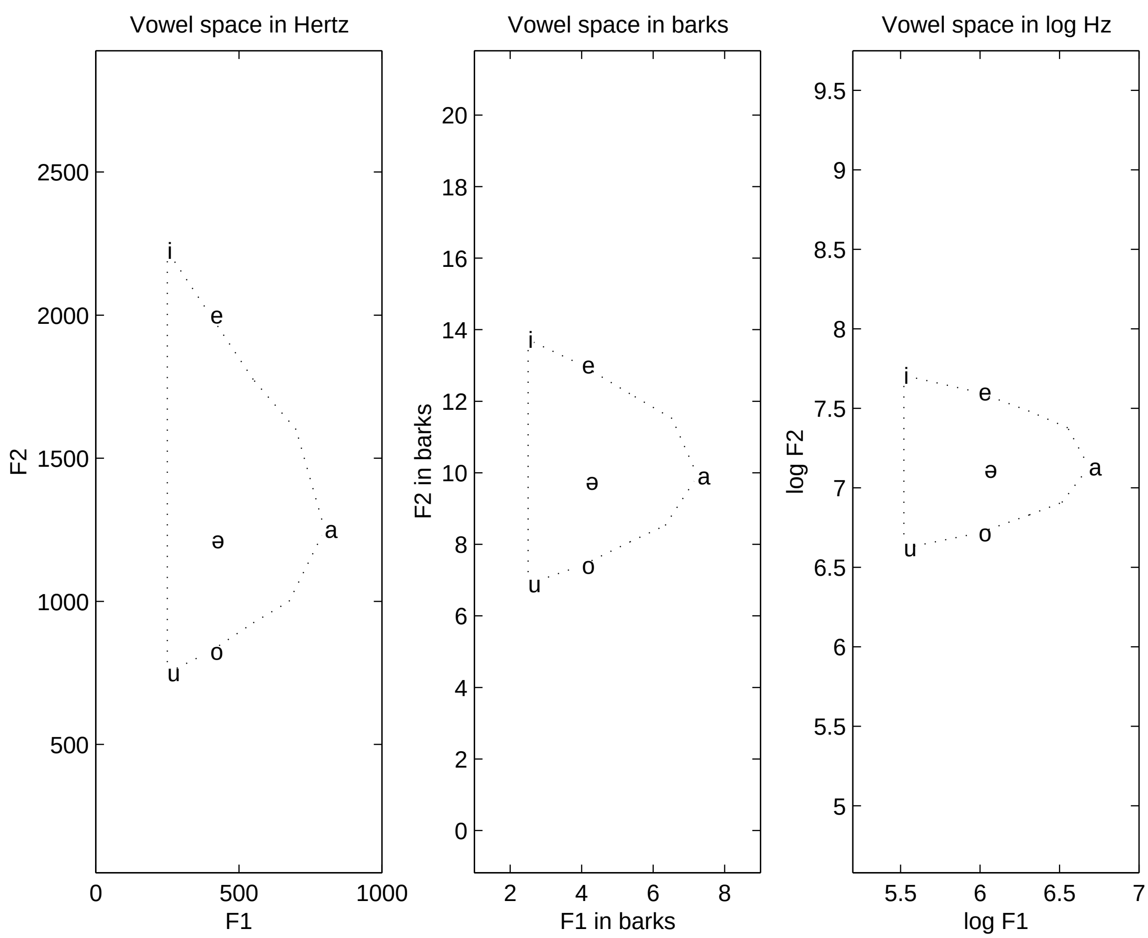

Figure 3. Representations of the vowel formant space with equally scaled F1 and F2 axes

If, however, we scale the *x* and *y* axes proportionally, we can see that these three representations of the vowel space are really quite different. Figure 3 shows the same graphs, but with the *x* and *y* axes proportional. The vowel space in Hertz is now seen to provide a relatively large range of values along the F2 dimension, compared to the F1 dimension, which, for a distance-oriented simulation, would imply more reliance upon F2 to maximize contrast. The Bark scale goes part of the way towards making F1 as important for total distance as F2, but there is still a larger range in the F2 dimension

[3] The points of each of the vowels were taken from LL72, and the log and Bark transformations were performed on those values.

(approximately 7 Barks) than in the F1 dimension (approximately 5 Barks). In a simulation that maximizes contrast by maximizing distance within the space, this provides more distance along the F2 dimension than along the F1 dimension for increasing contrast, which can be interpreted as a bias for high, non-peripheral locations. It is no accident that they predicted more of these than are found in the world's languages. The log scale, however, provides the F1 and F2 dimensions with the same sized range of possible values. If our goal is to treat F1 and F2 simply as two dimensions of contrast, the log scale will not favor one dimension over the other, simply in terms of the size of the range of values that can be taken

Given that we are going to use a two-dimensional Euclidean space, the simplest distance to use for finding ‘good' positions in the space is simple Euclidean distance. This is defined as the square root of the sum of the F1 distance squared and the F2 distance squared. This is what LL72 used[4], but a more complicated, non-linear distance metric was used in L86. The point we would like to make here is that, by changing the distance metric, they in essence changed the interpretation of the space. Keep in mind that distance in these vowel "spaces" is a metaphor for contrast, and if two points within the space that are relatively close together by standard measures are judged far apart by another metric, the effect is identical to a warping of the space. However, we are interested in the space itself: whether it represents an adequate formulation of vowel contrast, insofar as true “distance” in the space corresponds to discriminability. For this reason, we use the Euclidean distance within whatever vowel space we choose. If we wish to change our notion of contrast, we do so by changing the space explicitly. For example, when we begin to examine the issue of formant normalization, we will do so by changing the space that we stipulate, not by changing the distance metric.

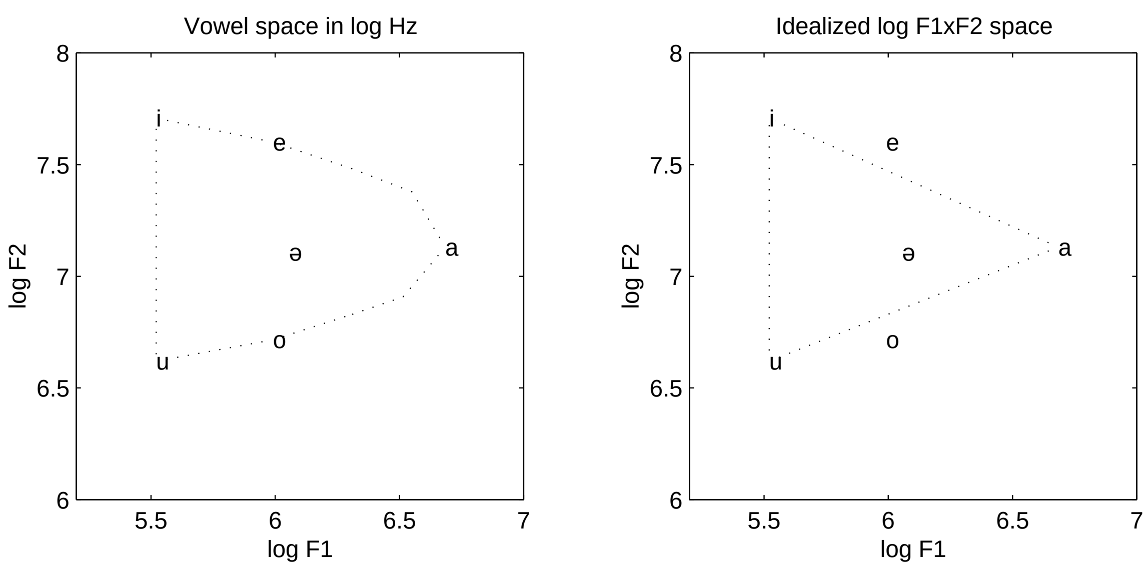

Figure 4. Log Hertz representation of the vowel space and an idealized Log Hertz vowel space

[4] In fact, rather than maximizing distance, all of the simulations, including ours, minimizes a distance-based energy function, which is one over the distance squared.

We will now give the rationale for adopting an idealized log Hertz space for our simulations. Figure 4 shows the log formant space and an idealized version of the same. Like the standard log Hertz formant space, the idealized version treats both dimensions (log F1 and log F2) as having an equivalent range. The treatment of the low space (around /a/) as angular rather than rounded is one way of building all quantal locations into the model. The three angles in the vowel space are such that, if a vowel gets pushed into the corner, it is unlikely to improve the total distance by coming back out again. This simple modification to the vowel space results in the points gravitating to these privileged locations and staying there.

As mentioned above, we will discuss this vowel space in terms of broad regions: front, back, and low peripheral, and non-peripheral. These regions were selected to allow for quick and easy comparison with the results of the UPSID survey. One could arguably include a larger number of regions, including, for example, central vs. high non-peripheral vowels or high front or back vs. non-high front or back. We limited it to these regions so that evaluating the simulation results with respect to the cross-linguistic tendencies would lead to a better understanding of the good and bad features of this approach. In other words, this regional configuration narrowly corresponds to the cross-linguistic generalizations that were mentioned above.

5. Simulations and results

One large piece of these vowel inventory simulations that has not been mentioned to this point is the search algorithm, which takes a random configuration and moves towards better ones. The basic idea is as follows: we provide the algorithm with a vowel inventory of a particular size, and see what it finds as the ‘best' configuration for that vowel inventory size. To illustrate how the algorithms work, let us consider the case where the vowel inventory size is five. The algorithms are given a delimited vowel space, and five randomly positioned vowels within that space. In LL72 and L86, the algorithm proceeded as follows: it chose one of the five vowels and moved it one step in the direction that increased the total distance between the vowels the most; it continued to move that same vowel until there was no move that stayed within the vowel space boundary and improved the total distance; it then selected the next vowel and did the same thing. This procedure iterated through the vowels until there was no movement of any vowel that could improve the total distance. The configuration that was found was then evaluated.

The previous simulations were run 15-30 years ago, and computing resources are much cheaper now than they were then, so we can afford to do more in our algorithm. Instead of moving one vowel as far as it can go before moving another, we evaluate 360 moves of the same distance (one degree difference) for each of the vowels in the configuration and choose the move that increases the distance the most from among all of the possible moves. Thus, for a five vowel system, we are evaluating 5*360 possible moves at each step of the algorithm. Once a move is made, we evaluate all possible subsequent moves. This procedure iterates until no moves improve the total distance.

While this algorithm is (perhaps) less prey to finding local rather than global maxima, it will still find local maxima. For this reason, we ran the algorithm 1000 times for each vowel inventory size with random starting points, and reported the percentage of trials for which a particular configuration is found. We also report the ‘energy' of the configuration, which is the inverse of the total distance squared.

Vowel Configuration					Languages Attested		Simulation Results		Previous Results	
Total Inventory	Peripheral			Non-Peripheral				Percent		
	Low	Front	Back		Number	Percent	Energy	Found	LL72	L86
3	1	1	1	0	17	89.5	2.1	100	*	*
3	1	2	0	0	1	5.25	5.4	0		
3	0	1	1	1	1	5.25	7.8	0		
4	1	2	1	0	14	56	8	69.8		
4	1	1	2	0	4	16	8.1	29.9		
4	1	1	1	1	7	28	8.2	0.3		
5	1	2	2	0	97	89	17.2	65.4	*	*
5	1	3	1	0	1	0.9	23.6	0		
5	1	2	1	1	5	4.6	17.9	20.9		
5	1	1	2	1	5	4.6	17.9	13.7		
5	1	1	1	2	1	0.9	31.3	0		
6	1	3	2	0	12	20	35.4	0		
6	1	2	3	0	4	6.7	35.3	0		
6	2	2	2	0	3	5	50.5	0		
6	1	2	2	1	41	68.3	30.1	100	*	*
7	1	3	3	0	27	56.2	58	0		
7	1	3	2	1	8	16.7	51.1	49.5		
7	1	2	3	1	0	0	50.8	44.9		
7	1	2	2	2	12	25	56.2	5.6	*	*
7	1	1	3	2	1	2.1	63.3	0		
8	1	4	3	0	3	15.8	93.5	0		
8	1	3	4	0	2	10.5	94.4	0		
8	1	3	3	1	8	42.1	76.3	72.3		
8	1	3	2	2	1	5.3	79.6	15.7	*	
8	1	2	3	2	0	0	79.3	12		
8	1	2	2	3	4	21	92.8	0		*
8	1	1	3	3	1	5.3	87.3	0		
9	1	4	4	0	7	29.1	142.6	0		
9	1	4	3	1	1	4.2	114.7	5.6		
9	1	3	4	1	0	0	115.4	4.2		*
9	2	3	3	1	1	4.2	151.5	0		
9	1	3	3	2	7	29.1	108.4	90.1		
9	1	4	2	2	4	16.7	118	0.1		
9	2	2	2	3	2	8.3	154.9	0	*	
9	1	2	3	3	1	4.2	118.9	0		
9	1	2	2	4	1	4.2	145.1	0		

Table 1. Vowel configurations: attested languages, ‘Quantal' simulation results, and previous results for inventories of size 3-9

Our first set of trials were carried out for vowel inventory sizes from 3 to 9, using the idealized log formant space shown above. Table 1 gives the results of these simulations, compared to both the attested languages for each configuration in the UPSID database, and the results from LL72 and L86. Those papers reported only a single configuration per size of vowel inventory, and their configuration is starred in the table. These first trials will be referred to as ‘Quantal' in future tables, since they depart most from previous work by encoding the quantal locations. We will discuss these results in detail in the next section.

Our second set of trials were an attempt to build some notion of formant normalization into our models of contrast. As mentioned above, vowels cannot be identified simply by first and second formant location, because the formant locations can and do differ widely from speaker to speaker. What must be recognized, then, is a formant pattern, some relationship between the formants. One way to encode this in our vowel space is to normalize the second dimension with respect to the first, which we do by dividing the second formant frequency by the first. Within such a vowel space, what is important in order to maximize contrast is not the distance between the second formant frequencies, but rather between the ratios of their second formants to their respective first formants. For example, suppose that there are two vowels such that the second formant is exactly twice the frequency of the first formant. In the ‘Quantal' vowel space that we used above, the distance between these two vowels would include some distance contributed by their first formant distance and some contributed by their second formant distance. In the new vowel space, with the normalized second dimension, which we will call ‘Normalized', there is no additional distance contributed by the second dimension, because we divide both second formants by their respective first formants, yielding 2 in both cases[5]. In this model, contrast between vowels is maintained as their F1 frequencies get closer together, provided the pattern that the F1 frequencies make with their F2 frequencies grow more distinct.

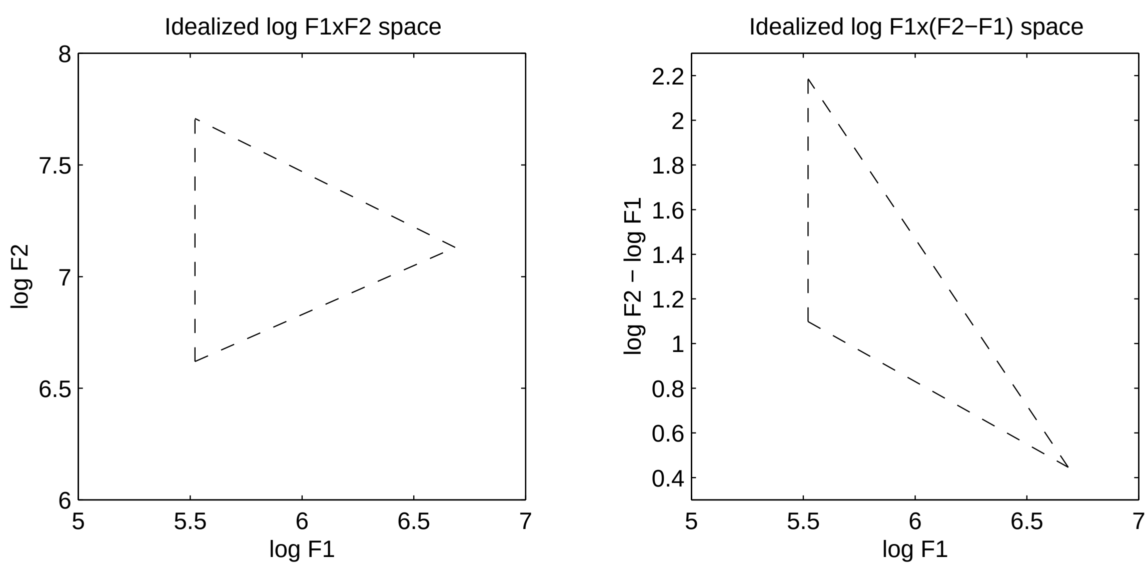

Figure 4. Two representations of the vowel space: what we call ‘Quantal' - logF1 x log F2; and what we call ‘Normalized' - log F1 x log (F2/F1)

[5] We take the log of this ratio in the simulations.

Most Attested Vowel Configuration						Simulation Configuration			
Total	Peripheral			Non-	Percent	This Paper (percent)		Previous (best)	
Inventory	Low	Front	Back	Peripheral	Attested	Quantal	Normalized	LL72	L86
3	1	1	1	0	89.5	100	100	*	*
4	1	2	1	0	56	69.8	100		
5	1	2	2	0	89	65.4	99.9	*	*
6	1	2	2	1	68.3	100	0	*	*
7	1	3	3	0	56.2	0	37.3		

*Table 2. Vowel configurations: most attested configurations (compiled from Schwartz et al., 1997), current simulation results, and previous simulation results. LL72 and L86 only reported 'best' configuration found, indicated with *.*

Figure 5 shows the two vowel spaces, 'Quantal' and 'Normalized', explicitly. One can see that the normalization stretches the periphery of the vowel space in a way that actually increases the range in the second dimension, but which provides more distance along the periphery for vowel contrast. In other words, this vowel space change, which we motivated through considerations of formant normalization, has the effect of favoring peripheral locations over non-peripheral locations, which is one of the cross-linguistic tendencies that we mentioned earlier in the paper. Also, the front periphery is more extended than the back periphery, which gives the asymmetric bias that was mentioned between front and back peripheral vowels. Simulations run with this vowel space ('Normalized') are shown in tables 2 and 3, compared with our 'Quantal' simulation, the LL72 and L86 simulations, and empirically observed configurations from Schwartz et al. (1997) and Crothers (1978). Table 2 shows, for vowel inventories of size 3-7, the percentage of times that the simulations find the most attested configurations, according to Schwartz et al. (1997), for that inventory size. We concentrated on these configurations, since they underline the differences between our two simulations and the previous results, and they contain the clearest failings of each approach. Table 3 labels configurations with the total number of vowels and the number of non-peripheral vowels (e.g. 4:1, meaning four total vowels, of which one is non-peripheral), and compares percentages of empirically attested configurations from both Crothers (1978) and Schwartz et al. (1997) with percentages reported in L86 and our own simulations.

6. Discussion of results

There are several points to make about these results. First, the 'Quantal' simulation improves on the previous results in the four vowel category, by finding all and only the attested configurations, and finding the most attested configuration most frequently, whereas neither the LL72 nor the L86 algorithms found matching configurations. The main problem in those algorithms was that they failed to keep vowels in the three corners. In addition, our simulation did successfully find, in contrast to the previous algorithms, common configurations even at the high inventories of 8 and 9. It did this by finding fewer non-peripheral vowels, which can be attributed to the fact that our vowel space favors neither F1 or F2 as primary dimensions of contrast.

The 'Quantal' simulation, however, fails fairly dramatically at the inventory size

Percentages reported from	Total:non-peripheral vowel configuration												
	3:0	3:1	4:0	4:1	5:0	5:1	5:2	6:0	6:1	6:2	7:0	7:1	7:2
Crothers (1978)	100	0	59	41	87	13	0	22	73	5	39	11	50
Schwartz et al. (1997)	95	5.3	72	28	90	9.2	0.9	32	68	0	56	17	27
L86 Simulation	76	24	38	62	10	80	0	0	68	32	0	12	82
Quantal Simulation	100	0	100	0.3	65	35	0	0	100	0	0	94	5.6
Normalized Simulation	100	0	100	0	100	0	0	100	0	0	48	52	0

Table 3. Percentage of total and non-peripheral vowel configurations (T:NP) from attested configurations (Crothers, 1978 and Schwartz et al., 1997) and from simulations. Crothers was the empirical gold standard cited in L86.

of seven, as do both the LL72 and L86 simulations. There are two distinct failures of the ‘Quantal' simulation: first, the simulation never finds the most common configuration, which has an equal number of front and back peripheral vowels and no non-peripheral vowels; also, it contains no bias towards front vowels when there is an asymmetry, i.e. it finds a configuration with 2 front peripheral and 3 back peripheral nearly as frequently as a 3 front and 2 back configuration, yet the former configuration is unattested for vowel inventories of size seven, according to Schwartz et al. (1997). Hence, if the claim is that contrastiveness is the critical force in shaping likely vowel inventories, then this model is not capturing all operative dimensions of contrastiveness.

The ‘Normalized' simulation, however, with its built-in periphery bias, and a further bias towards the front periphery versus the back periphery, finds the most common configuration for an inventory of size seven in more than a third of the trials. Furthermore, the remainder of the trials are asymmetric with respect to the number of front and back vowels on the periphery, and they all have more front than back peripheral vowels. Hence, we do have a model of vowel contrast that can at least suggest an explanation for the seven vowel tendencies.

Despite these successes of the ‘Normalized' simulations, they do make a number of poor predictions, generally also because of the bias for the periphery. Six vowel inventories most frequently have a non-peripheral vowel, which the ‘Normalized' simulation fails to predict. Attested seven vowel inventories have two peripheral vowels more frequently than they have a single peripheral vowel, and the ‘Normalized' simulation never finds such a configuration. It seems that both of our simulations make some good and some bad predictions.

While most of the LL72 and L86 results were reported simply as the ‘best' configuration found by the simulation, L86 also presented a table that grouped configurations by their total inventory size (from 3-7 vowels) and the number of non-peripheral vowels in the configuration, and reported the percentage of their top 50 configurations for that inventory size that had that number of non-peripheral vowels. Table 3 reproduces their table, and adds the Schwartz et al. (1997) empirical data and our simulation data. The preference for positing non-peripheral vowels in the L86 simulations can be seen very clearly in this table: at inventories of size four and five, they reverse the empirical tendencies by finding peripheral vowels more often than not. Our simulations have the opposite bias, predicting non-peripheral vowels in some cases less frequently than they actually occur.

There are a couple of general issues that can be addressed at this point, which bear on the failure of simulations of this sort to fit the distribution of attested languages. First, there is no indication in the UPSID survey regarding the relationships between languages that share configurations. Related languages may share a vowel configuration by virtue of a common ancestor language, rather than an independent convergence on that configuration. Thus, some of the distribution may be somewhat explained by historical accident rather than by some systemic bias. Also, vowel shifts do not begin from random starting points, which may further obscure systemic biases. Furthermore, there are far fewer languages with higher vowel inventories, so there may be a difficulty in inferring the 'true' distributional bias from this sample. That said, there are enough languages that fall in some of our more troubling configurations, that we can probably still take the failure of our simulation to find these as a problem.

More crucially, there some fundamental questions about the nature of global contrast and the way that this simulation is carried out. First, it is not clear that 'maximal' contrast, or even locally maximal contrast, is necessarily better perceptually than some 'sufficient' level of contrast. Perhaps contrast maintenance is active, but only really governs configurations when vowels become too close. Given the amount of variation among attested configurations, this is almost certainly true in some respect. Our energy function, by using the square of the distance, helps somewhat in this respect, since the energy becomes very small much faster than the simple reciprocal of the distance, effectively shrinking the scope of a vowels influence within the space. It remains an open question, however, as to whether the 'goodness' of a configuration can continue to improve beyond some point where it is 'good enough'.

Secondly, it is also not clear that distance should be maximized globally. Consider a seven vowel system with two positions for a vowel that are equi-distant from all of its immediate neighbors, but with one of the positions farther from vowels that are already quite distant in the space. In this case, it seems that the local distance is all that is truly relevant for contrast preservation. What makes this a difficult perspective to test is the algorithm for vowel dispersion: in a random configuration, the local interactions may not be sufficient to guide the vowels to stable states. It remains to be fully explored.

Finally, given that the different simulations seem to make good predictions some of the time, one might think that perhaps they can be combined into a single model that makes predictions that better match the empirical distributions. While this may be possible, it departs somewhat from the spirit of the enterprise, which is really intended to ask the questions: what counts as perceptual contrast between vowels, and can some well-motivated notion of contrast explain why certain configurations tend to occur more frequently than others. We have proposed two alternate notions of contrast, one based on formant frequency points, and one based on formant patterns, and they each seem to capture some of the empirical distribution, but not all. To arbitrarily model the process as a composite of the two seems unmotivated, and not particularly informative with respect to the questions at hand.

In sum, these simulations provide more food-for-thought for the question of why certain vowel inventories are more common, as well as for the question of what constitutes contrast in vowel systems. We have demonstrated that formant distance alone is insufficient to account for all of the empirical tendencies, and that certain simple improvements to a model of the vowel space (e.g. including preferred quantal locations), which enhance the degree to which actual contrastiveness is represented in the model, can make a large improvement in the degree to which these tendencies can be explained by the model. Furthermore, we have suggested that there are other well-motivated ways to think about vowel contrast that move beyond formant frequency locations, in ways that are motivated by certain uncontroversial features of vowel perception. Vowel perception is not an easy problem, and it is unsurprising that the demands of this perceptual process might exert some influence on the shape of vowel inventories.

7. Conclusion

To conclude, we have presented a re-implementation of the widely-cited vowel simulations of Liljencrants and Lindblom, first with some simplifications and some small, well-motivated, enhancements, and also with some more dramatic changes. We have found that good predictions, above and beyond those found previously, did arise in each of our simulations, but some problems remained. In particular, it seems that we can explain, at least to a certain extent, tendencies in previously problematic inventory sizes using contrast, but that a single definition does not seem to suffice in all cases.

This paper's contribution is twofold. First, it cleared up certain simple failings of the earlier simulations by slightly modifying the stipulated vowel space. These simulations ('Quantal') are relatively straightforward extensions of the previous work. Second, it suggested a new way of thinking about vowel contrast, in a way that departs from simple frequency comparisons to include fundamental aspects of vowel perception. These simulations ('Normalized') are a large departure from the previous work, and, while the results can really only be called suggestive, they do point to a perspective that may be quite viable in this area.

8. Acknowledgements

Thanks to Molly Homer for helpful comments and criticisms, and to Phil Lieberman for general discussion on this topic. This work was partially supported under NSF IGERT grant #DGE-9870676.

References

Crothers, J. 1978. Typology and universals of vowel systems. In J.H. Greenberg, C.A. Ferguson, and E.A. Moravcsik, editors, Universals of human language, Vol. 2: Phonology. Stanford University Press, Palo Alto, CA.

Flemming, E. 1996. Evidence for constraints on contrast: The dispersion theory of contrast. In C.S. Hsu, editor, UCLA Working Papers in Phonology 1. University of California, Los Angeles.

Hayes, B. 1999. Phonetically-driven phonology: The role of optimality theory and inductive grounding. In M. Darnell, E. Moravscik, and M. Noonan, editors, Functionalism and Formalism in Linguistics, Volume I: General Papers. John Benjamins, Amsterdam.

Liljencrants, J. and B. Lindblom. 1972. Numerical simulation of vowel quality systems: the role of perceptual contrast. Language, 48(4):839--862.

Lindblom, B. 1986. Phonetic universals in vowel systems. In J. Ohala and J. Jaeger, editors, Experimental Phonology. Academic Press, New York.

Nearey, T. 1978. Phonetic Feature Systems for Vowels. Indiana University Linguistics Club, Bloomington, Indiana.

Schwartz, J., L. Boë, N. Vallé, and C. Abry. 1997. Major trends in vowel system inventories. Journal of Phonetics, 25:233--253.

Silverman, D. 1998. Alveolar stops in American English, and the nature of allophony. In P.N. Tamanji and K. Kusumoto, editors, NELS 28. GLSA, University of Massachusetts, Amherst.

Stevens, K.N. 1972. The quantal nature of speech: Evidence from articulatory-acoustic data. In P.B. Denes and E.E. David Jr., editors, Human communication: A unified view. McGraw-Hill, New York, pages 51\66.

Department of Cognitive and Linguistic Sciences
Box 1978
Brown University
Providence, RI 02912

brian-roark@brown.edu