24.910 Topics in Linguistic Theory: Laboratory Phonology
Spring 2007

24.910
Laboratory Phonology
The Theory of Adaptive Dispersion

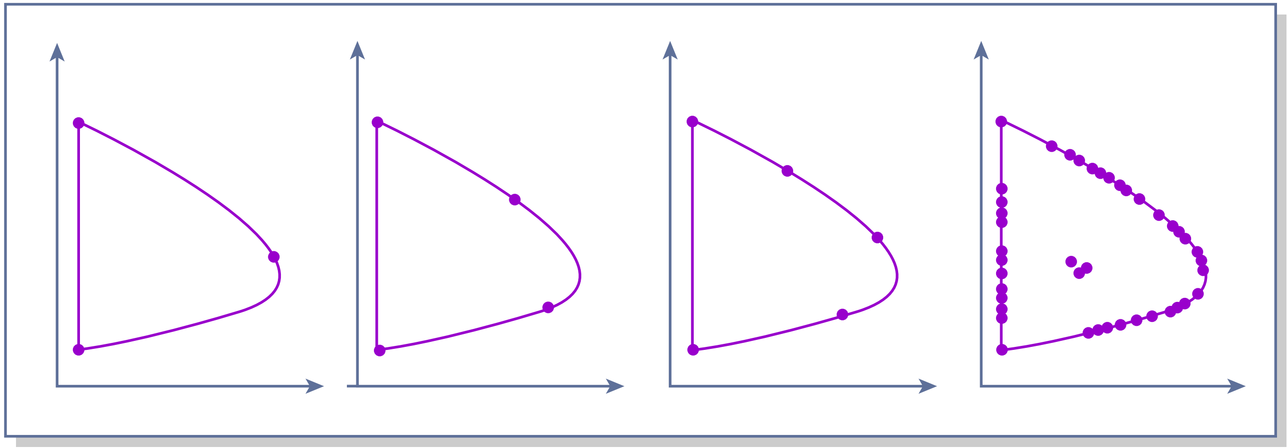


Image by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Readings for next week:

- Steriade (1999), pp. 1-21
- Wright (2004).

Assignment:

- Waveform editing

Lindblom's Theory of Adaptive Dispersion

- Common vowel inventories:

i u	i u	i u
		e o
	e o	ɛ ɔ
a	a	a
Arabic, Nyangumata, Aleut, etc.	Spanish, Swahili, Cherokee, etc.	Italian, Yoruba, Tunica, etc.

- Unattested vowel inventories:

i	i ɯ	i u
		ɪ ʊ
e	e ɤ	
a	a	

Lindblom’s Theory of Adaptive Dispersion

- Try to explain why vowel systems are the way they are.
- Observation: vowels in an inventory tend to be evenly dispersed through the vowel space (cf. Disner 1984).
- Hypothesis: this facilitates efficient communication by minimizing the likelihood of confusing vowels.

i u	i u	i u
		e o
	e o	
		ɛ ɔ
a	a	a

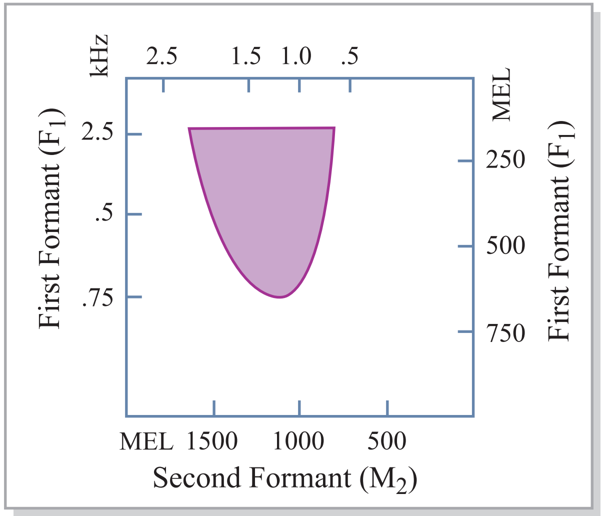


Figure by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Lindblom’s Theory of Adaptive Dispersion

- Try to explain why vowel systems are the way they are.
- Observation: vowels in an inventory tend to be evenly dispersed through the vowel space (cf. Disner 1984).
- Hypothesis: this facilitates efficient communication by minimizing the likelihood of confusing vowels.

- Vowels that are closer in the perceptual space are more easily confused.
- Confusions between contrasting sounds impair communication.
- So contrasting vowels should be as far apart as possible (dispersion).

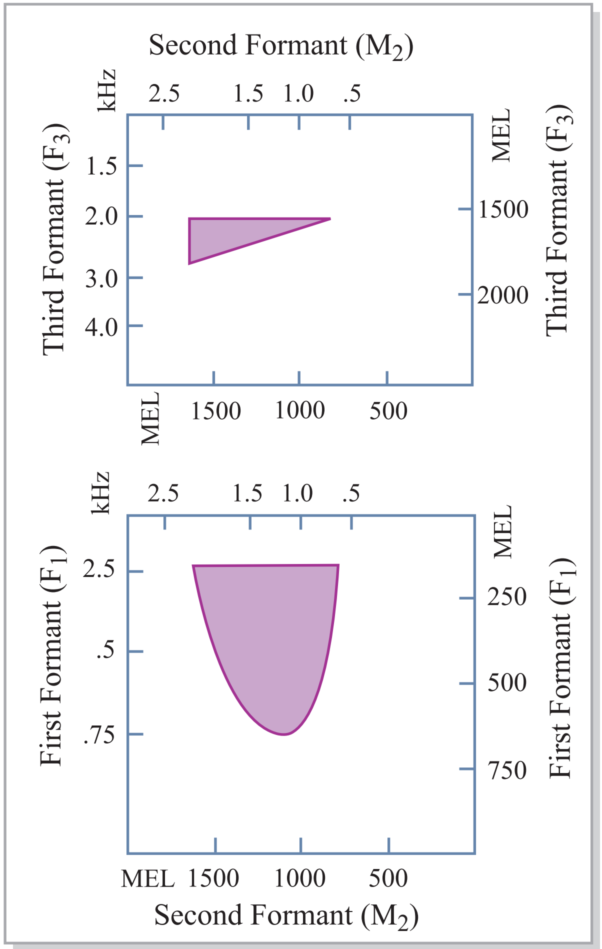
Liljencrants & Lindblom (1972)

Approach to exploring dispersion hypothesis:

- Modeling
- Simulation
- Comparison of simulation results to impressionistic descriptions of a large sample of vowel inventories.

Liljencrants and Lindblom (1972)

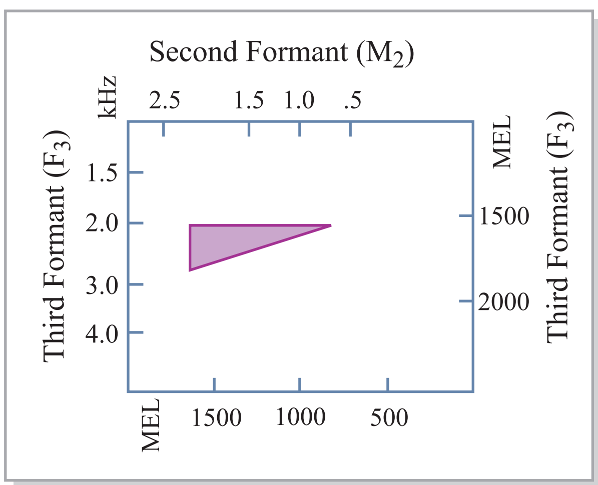
- The role of perceptual contrast in predicting vowel inventories.
- The perceptual space of articulatorily possible vowels:

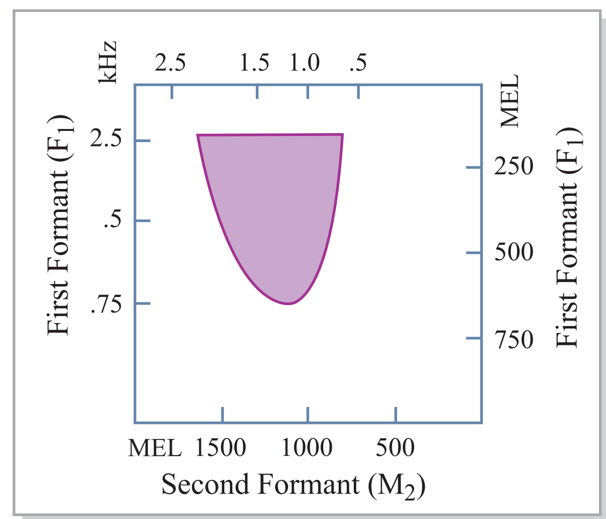


Figures by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

The vowel space

- Why does the vowel space look like this?
  - ➢ Why do the dimensions correspond to formant frequencies?
  - ➢ Why just the first 2-3 formant frequencies?
  - ➢ Why does the F1-F2 space have this shape?





Figures by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Why do the perceptual dimensions of vowel quality correspond to formant frequencies?

(cf. Pierrehumbert 2000)

Production - we can control formant frequencies.

- Given that vowels are produced with a relatively open vocal tract, the properties of these sounds that we can manipulate most easily are:
  - f0 (pitch) - a source property. The basis for tone contrasts.
  - formants - filter property - the resonant frequencies of the vocal tract.
- Bandwidths and formant intensities generally covary with formant frequencies (Fant 1956).
- Varying bandwiths independently would involve changing the stiffness of the vocal tract walls, or the mode of vocal fold vibration. (NB nasalization affects formant bandwidths).

Why do the perceptual dimensions of vowel quality correspond to formant frequencies?

Perception - we can perceive formant peaks.

- f0 is (usually) much lower than formant frequencies.
  - Resonant frequencies are well represented as peaks in the ouput spectrum.
  - Exception: soprano singing.
- Formant peaks are more robustly perceptible than ‘valleys’ because they can rise above background noise.

Why do the perceptual dimensions of vowel quality correspond to F1, F2 (&F3)?

Higher formants are not important in vowel quality because they are insufficiently perceptible (especially in noise).

- There is less energy in the voice source at higher frequencies.
- Our ears are less sensitive to higher frequencies.

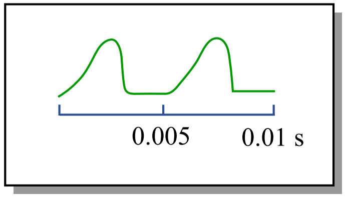


Figure by MIT OpenCourseWare.

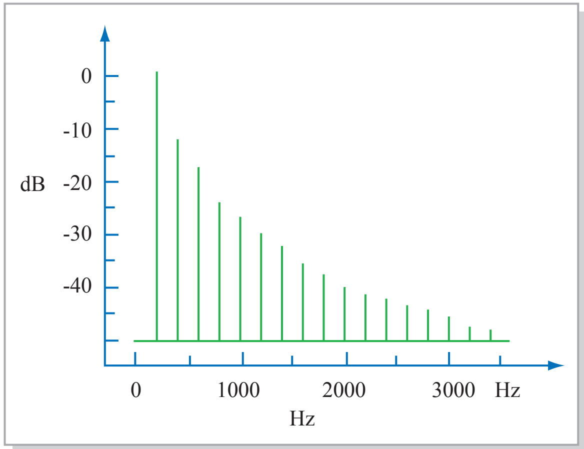


Figure by MIT OpenCourseWare.

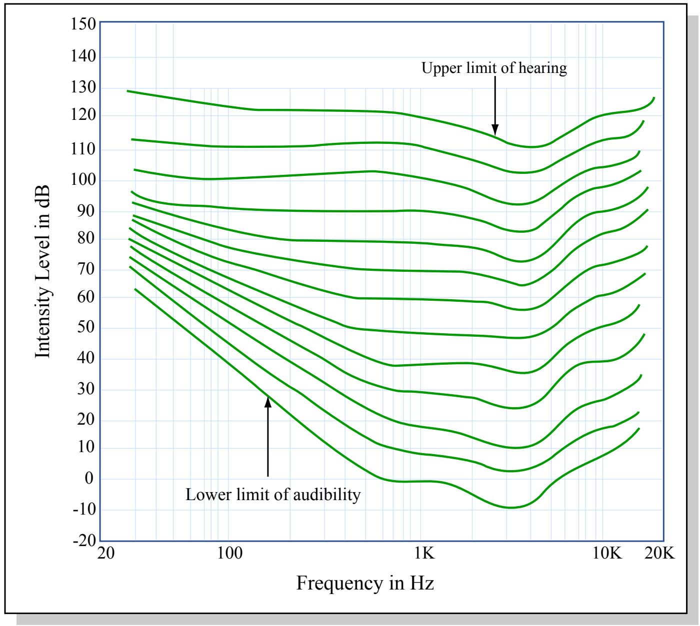
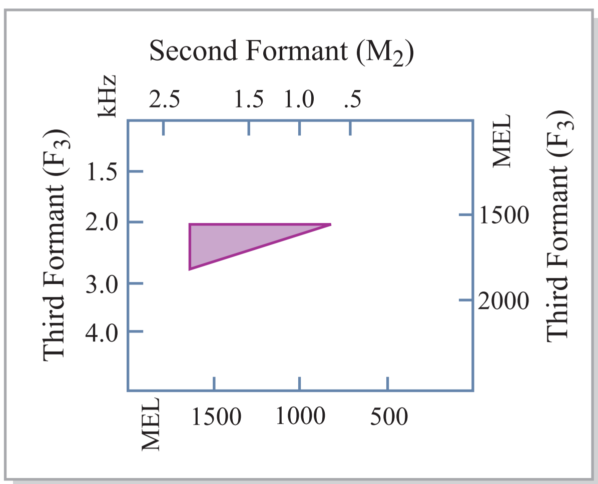
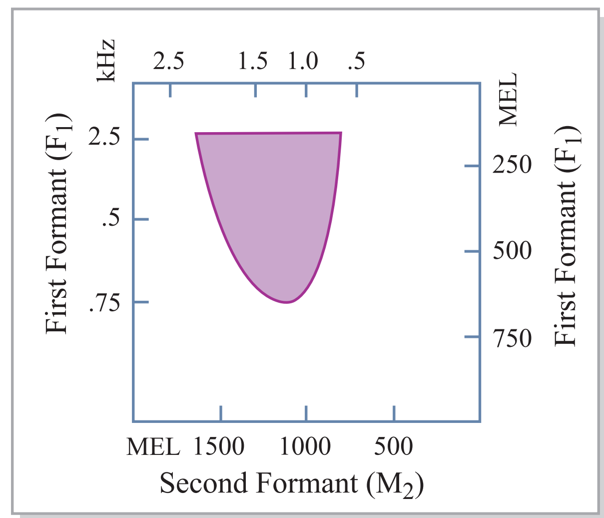


Figure by MIT OpenCourseWare.

The vowel space

- Why does the range of possible F2 values taper as F1 increases?
- How do you achieve maximum and minimum F1?
- How do you achieve maximum and minimum F2?





Figures by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Liljencrants and Lindblom (1972)

- Perceptual distinctiveness of contrast between V_i and V_j: distance between vowels in perceptual vowel space

$$r_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}$$

where x_n is F2 of V_n in mel

y_n is F1 of V_n in mel

- Maximize distinctiveness: select N vowels so as to minimize E

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{r_{ij}^2}$$

Liljencrants and Lindblom (1972)

- Prediction: vowel inventories with a given number of vowels should arrange those vowels so as to minimize E.
- What are those predicted vowel arrangements?
  - Optimization problem: For N vowels, find F1, F2' values that minimize E (objective function).
  - Large search space, many local minima.

$$E = \sum_{i=1}^{n-1} \sum_{j=0}^{i-1} \frac{1}{r_{ij}^2}$$

Minimizing E - stochastic search

- Start with vowels arranged in a circle near the center of the vowel space. (Random arrangement might be better?)
- Pick a vowel at random.
  - Try small movements of that vowel in 6 directions (within the vowel space)
  - Select the direction that results in greatest reduction in E.
  - Move vowel in that direction until E stops decreasing, or a boundary is reached.
- Repeat for all vowels.
- Cycle through the vowels until no further reduction in E can be achieved.
- Should be repeated multiple times, preferably with different starting configurations.
- More sophisticated search strategies are possible, e.g. simulated annealing or more sophisticated procedures for identifying best change at each stage.

- Predicted optimal inventories
- Reasonable approximations to typical 3 and 5 vowel inventories are derived.
- Preference for [i, a, u] is derived.
- Problem: Too many high, non-peripheral vowels.
- Not enough mid non-peripheral vowels.

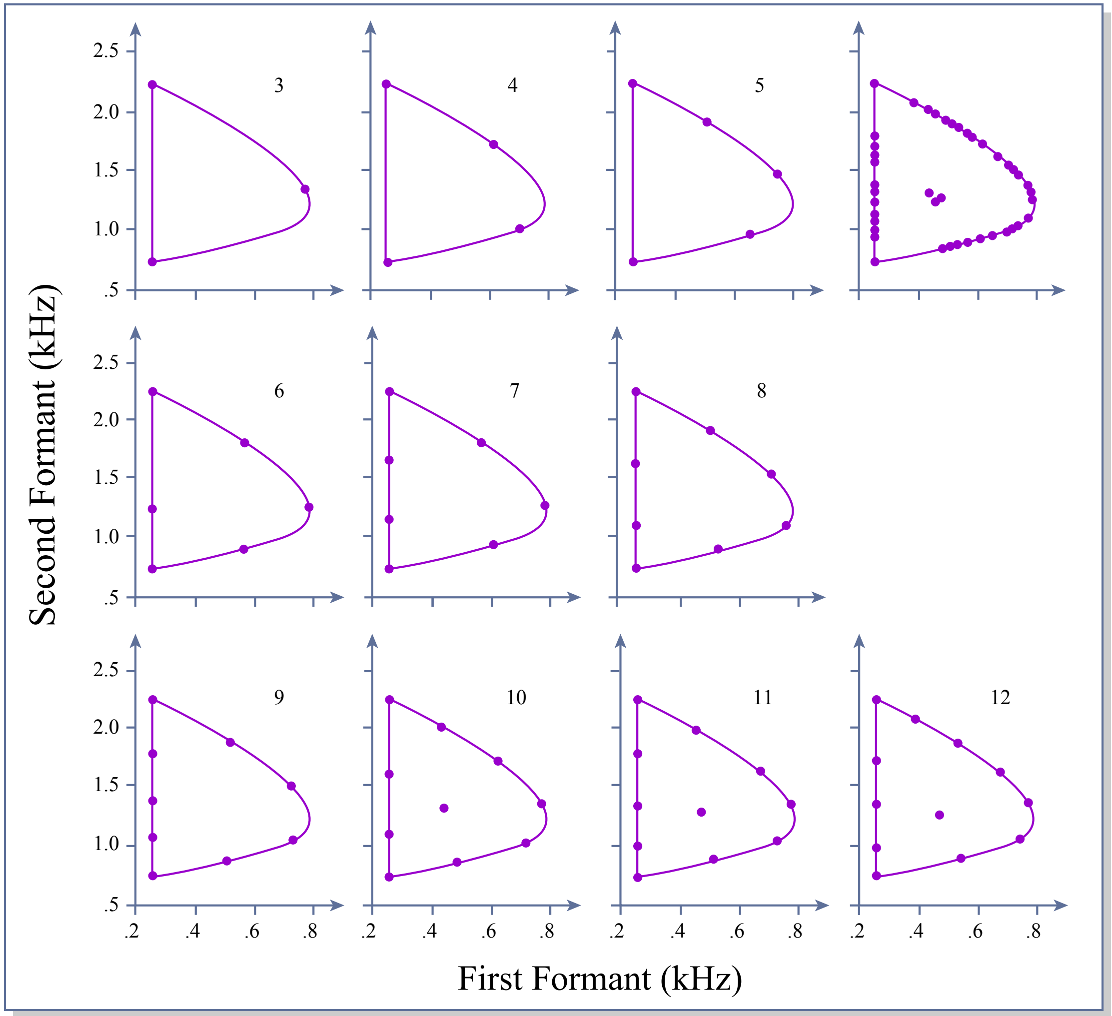


Figure by MIT OpenCourseWare. Adapted from Liljencrants, Johan, and Bjorn Lindblom. "Numerical Simulation of Vowel Quality Systems: The Role of Perceptual Contrast." *Language* 48, no. 4 (December 1972): 839-862.

Too many high non-peripheral vowels

- All inventories larger than 5 are predicted to contain one or more high vowels between [i] and [u], e.g. [y, ɨ, ɯ].
- E.g. prediction for 7 vowels (unattested):

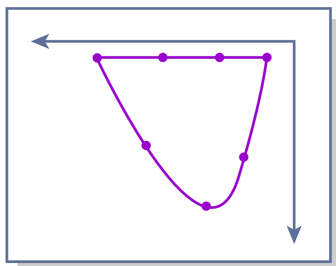


Figure by MIT OpenCourseWare.

i	ü/ʉ		ɨ	u
ɛ				ɔ
		a		

Figure by MIT OpenCourseWare.

- Common 7 vowel inventories:

i	y	u		i		u
e	ø	o		e		o
				ɛ		ɔ
	a				a	

Figure by MIT OpenCourseWare.

Too many high non-peripheral vowels

- The excess of central vowels arise because measuring distinctiveness in terms of distance in formant space gives too much weight to differences in F2.
  - In general, languages have more F1 contrasts than F2 contrasts.
- Why are F1 differences more distinct than F2 differences?
- One factor: auditory sensitivity to frequency (next slide).
  - But L&L already took this into account - mel scaled formant frequencies.

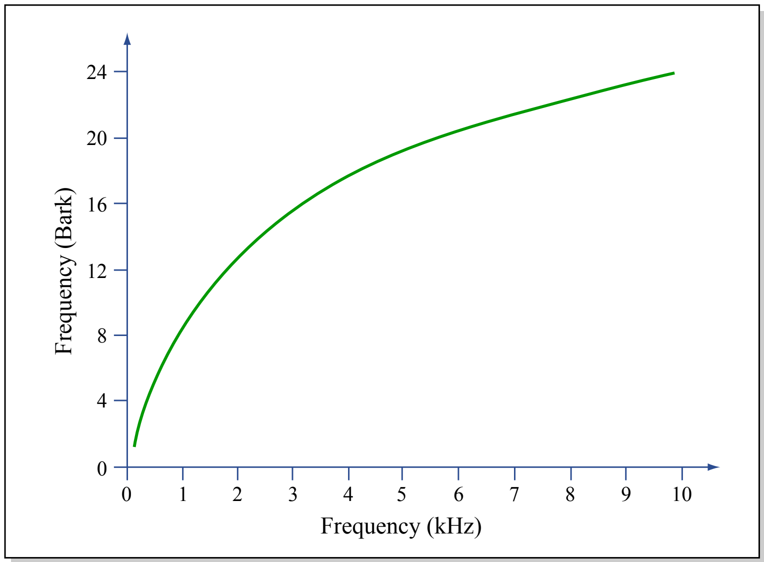
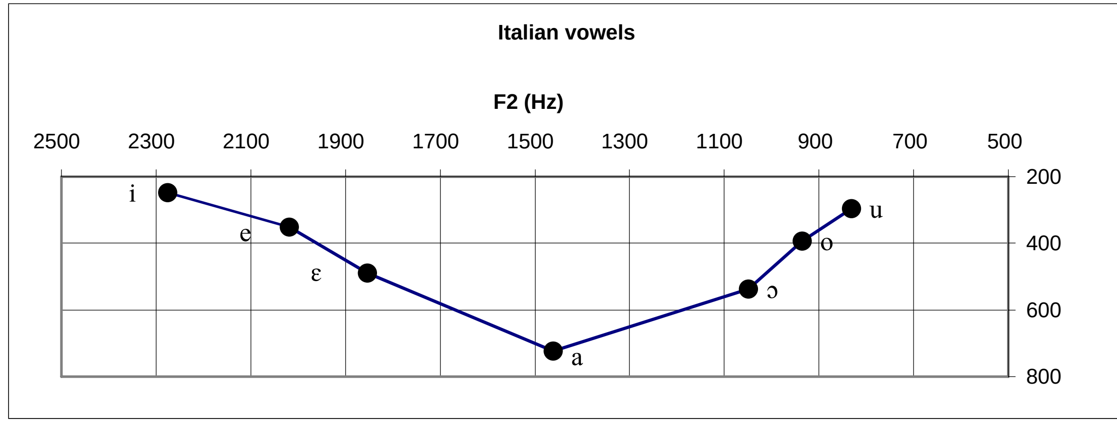
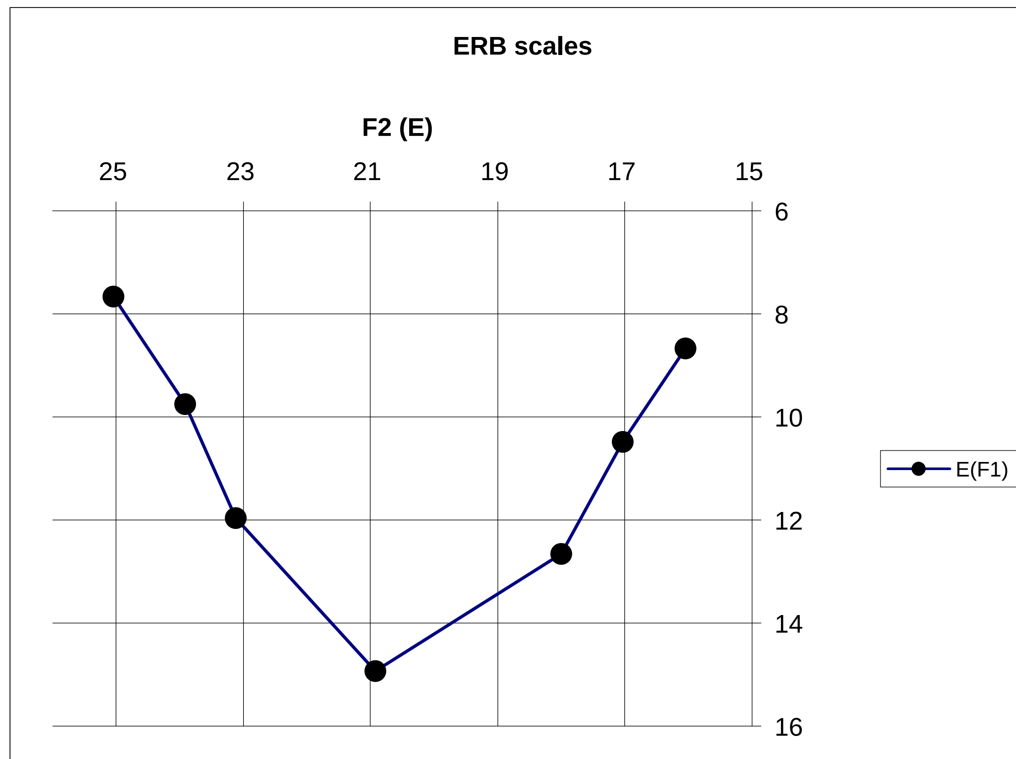


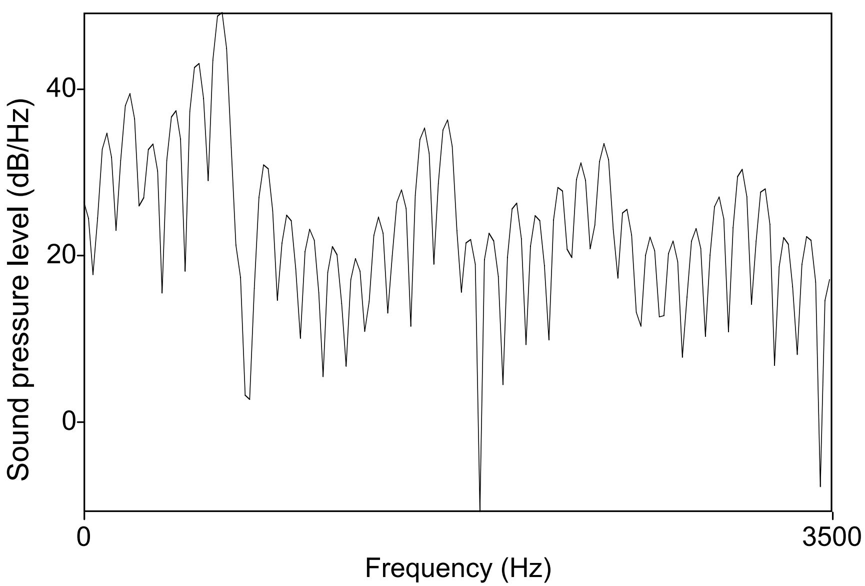
Figure by MIT OpenCourseWare. Adapted from Johnson, Keith. *Acoustic and Auditory Phonetics*. Malden, MA: Blackwell Publishers, 1997. ISBN: 9780631188483.

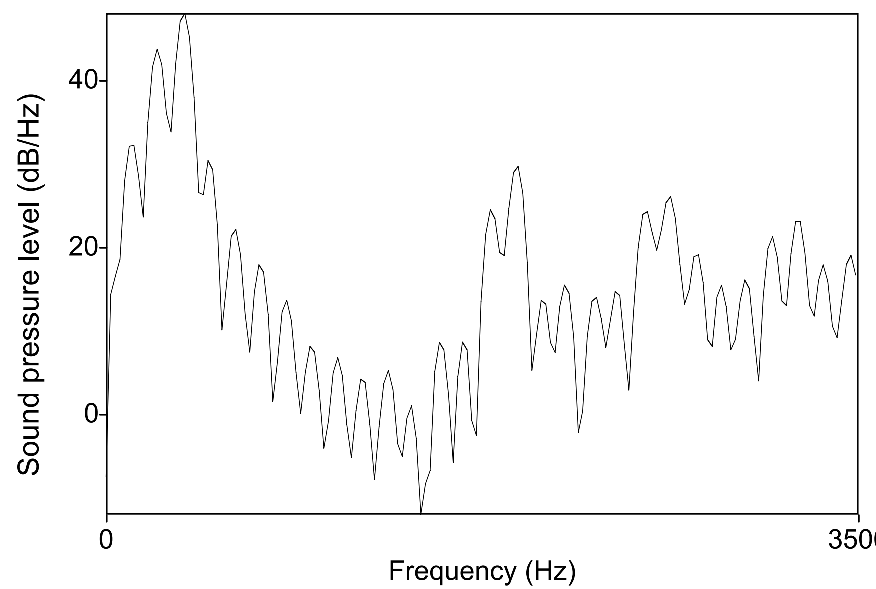




Too many high non-peripheral vowels

- Recent work by Diehl, Lindblom and Creeger (2003) suggests that the greater perceptual significance of F1 probably follows from the higher intensity of F1 relative to F2.
  - F1 should be more salient auditorily and more robust to noise.





Too many high non-peripheral vowels

- New simulations of 7 vowel system by Diehl, Lindblom and Creeger (2003)
  - incorporate background noise
  - perceptual distance is calculated as difference between auditory spectra.

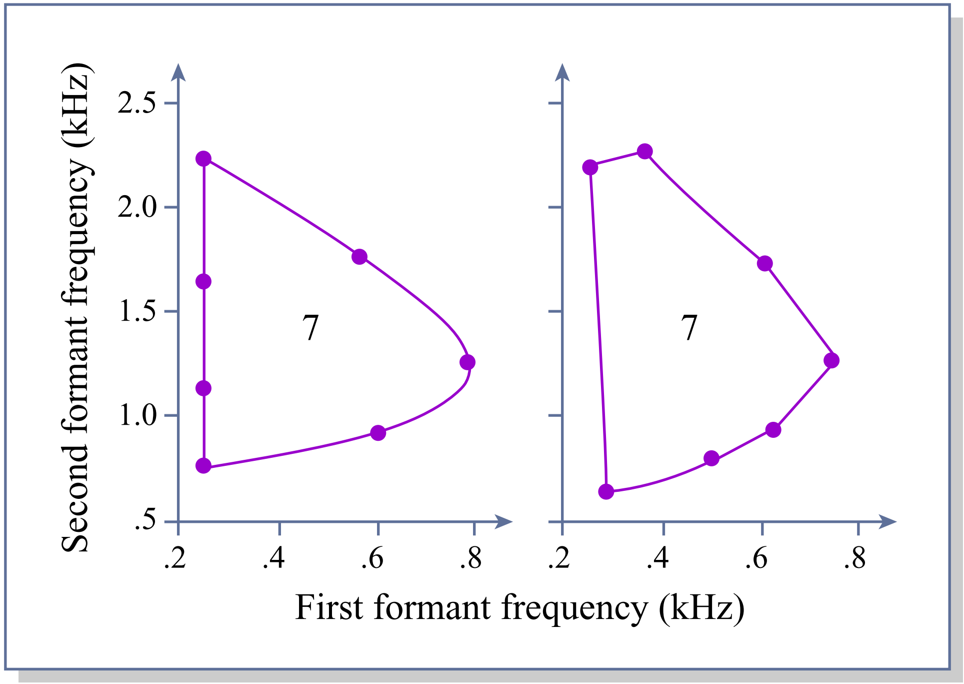


Figure by MIT OpenCourseWare. Adatped from Diehl, R. L., B. Lindblom, and C. P. Creeger. "Increasing Realism of Auditory Representations Yields Further Insights Into Vowel Phonetics." *Proceedings of the 15th International Congress of Phonetic Sciences.* Vol. 2. Adelaide, Australia: Causal Publications, 2003, pp.1381-1384.

The ‘corner’ vowels [i, a, u]

- Considerations of formant intensity might also help to account for some exceptions to the generalization that every language includes the ‘corner’ vowels [i, a, u].
- L&L predict that this should be the case, and most languages do include all three, but a number of languages lack [u]:
  - [i, a, o], e.g. Piraha, Axeninca Campa
  - [i, e, a, o], e.g. Navajo, Klamath
  - [i, e, a, o, ɯ], e.g. Tokyo Japanese
- In general F1 is more intense where it is higher, and this also raises the intensity of all higher formants. In [u], both F1 and F2 are low, resulting in a low intensity vowel, with low intensity F2.

Too few interior vowels

- When an inventory has mid vowels [e, o] and front rounded vowel [y], it often has mid front [ø] as well (Finnish, German, French, etc)
- L&L predict that interior vowels only appear with 10 or more vowels.
- The absence of interior vowels [ə, ø] is a result of the way in which overall distinctiveness is calculated.
- Each vowel contributes to *E* based on its distance from every other vowel.
- Interior vowels have a high cost because they are relatively close to all the peripheral vowels.
- Perhaps the measure of distinctiveness, *E*, can be improved on.

i	y		u	i	ɨ	u
e	ø		o	e	ə	o
		a			a	

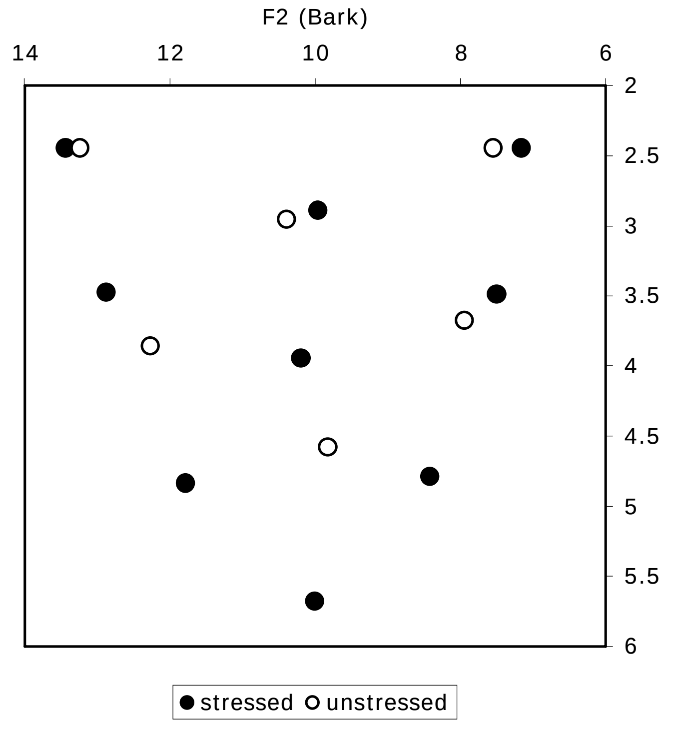
Figure by MIT OpenCourseWare.

Alternative measures of distinctiveness

- L&L's measure E is based on an analogy to dispersion of charged particles - it is not derived from anything based on vowel perception.
- It has the important property that distinctiveness 'cost' increases more rapidly as two vowels become closer - $1/r_{ij}^2$
  - I.e. vowels are only likely to be confused if they are quite similar. Likelihood of confusion drops of quickly as distance increases.
- But perhaps $1/r_{ij}^2$ doesn't drop off quickly enough - the lack of interior vowels results from giving too much weight to vowel pairs that are not very close.
- An alternative (Flemming 2005): only consider the closest pair of vowels in the inventory.
- Compromise (to be explored): $1/r_{ij}^n$, $n > 2$.

Alternative measures of distinctiveness

- Maximize the minimum distance (Flemming 2005)



Problems with Adaptive Dispersion

- Specific instantiations of the model have made specific incorrect predictions (but some of the broad predictions are correct and models are improving).
- The model answers an inobvious question: 'Given N vowels, what should they be?' - what determines the size of inventories?
- TAD predicts a single best inventory for each inventory size. Why would languages have sub-optimal inventories?
  - The unattested inventories shown earlier are obviously very porrly dispersed, but there are a variety of attested inventory patterns for any given number of vowels.

Extending Adaptive Dispersion

- If perceptual distinctiveness is important in shaping vowel inventories, then it should play a similar role in shaping consonant inventories.
- It is harder to develop quantitative models in this area because it is less clear what the perceptual dimensions are.
  - Especially because many consonants cannot be treated as static, e.g. stops.
  - Note that this is an issue for vowels also - how do diphthongs and vowel duration contrasts fit into the model?