Phonological features and sub-categorical structure in speech perception.

Noah Silbert & Kenneth de Jong
nosilber@indiana.edu kdejong@indiana.edu
Linguistics & Cognitive Science
Indiana University
www.iub.edu/~lsl PhonologyFest 2006

Phonological Feature Theory

Distinctive (contrastive) features typically

- are binary, unordered nominal variables
- are based on articulation
- are given ‘axiomatically’ as a foundation
- group speech sounds in overlapping categories

e.g.,

	manner	
	place	place
voicing	[p] [t] [k]	[f] [θ] [s]
	[b] [d] [g]	[v] [ð] [z]

Phonological Feature Theory

Evidence for (the utility of) distinctive features

- speech sounds (phones) arrayed in 'parallel' across a wide variety of languages

e.g., [p] [t] [k] [f] [θ] [s]
[b] [d] [g] [v] [ð] [z]

- reasonably accurate description of diachronic and synchronic sound correspondences in a wide variety of languages.
- However, features are quite coarse.

Implications of Feature Theory

Feature Equivalence (FE)

- Contrast on any given feature is as distinctive as contrast on any other feature.
- e.g., [p]-[b] and [p]-[f] are equally contrastive.

Absence of Feature Interaction (FI)

- Within a phone, any one feature specification is irrelevant to any other feature specification.
- e.g., [p]-[b] and [f]-[v] are equivalent
- features should not interact with syllable position
- e.g., [pa]-[ba] and [ap]-[ab] are equivalent

Speech Perception and FE, FI

FE and FI have been addressed more or less directly in the speech perception literature

- Miller & Nicely (1955): voicing more robust than place to noise and low-pass filter degradation (!FE)
- Wang & Bilger (1973): voicing more important than most place and all manner features (!FE)
- Goldstein (1980), Shepard (1972), Shepard & Arabie (1979), Soli & Arabie (1979), Klatt (1968), Soli, Arabie, & Phipps (1986): abundant sub-categorical structure across tasks, conditions (FI)
- Above studies all based on group data.

A Reanalysis of Some Recent Data

Cutler, Weber, Smits, & Cooper (2004)

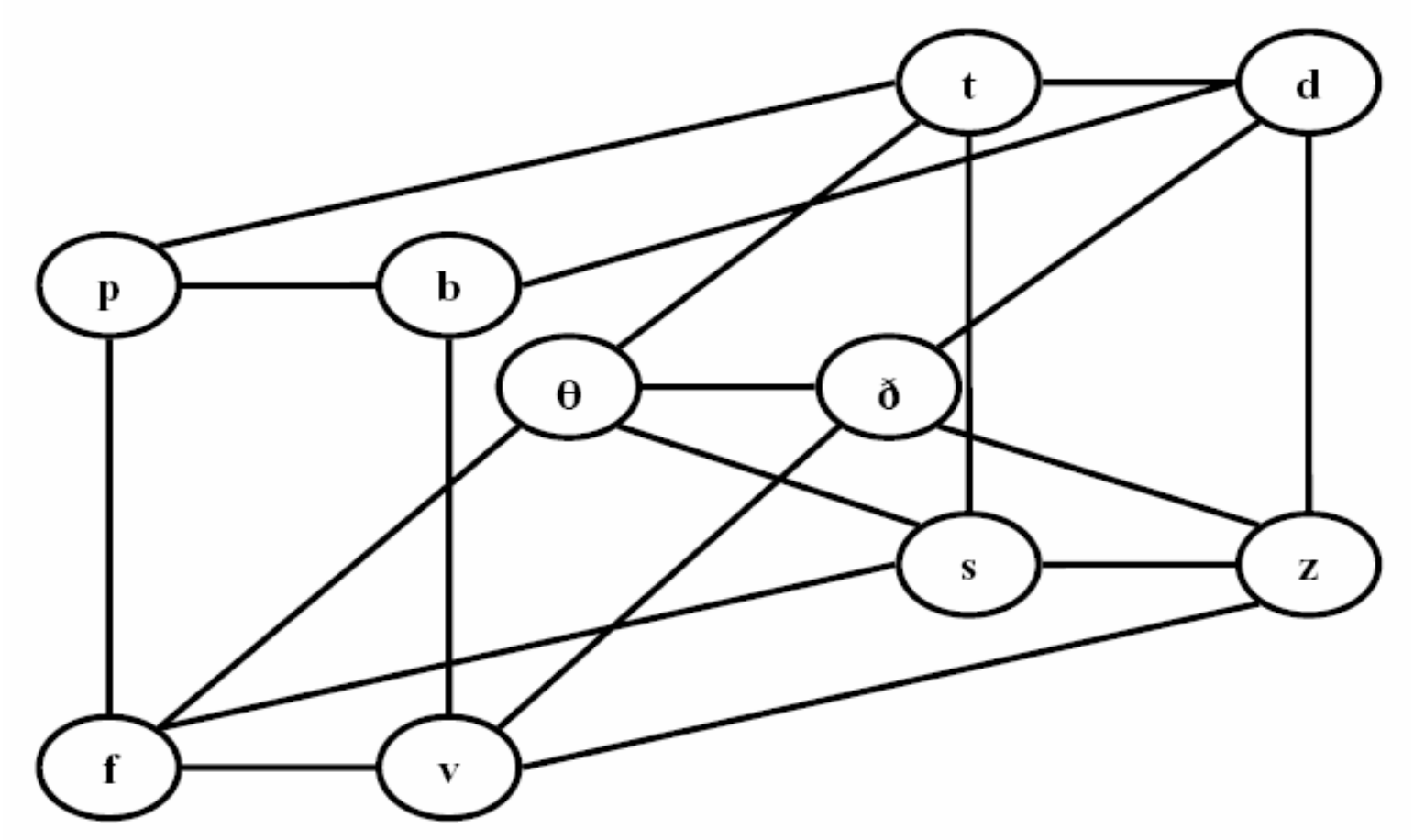
- consonant ID task
- 22 English consonants
- 16 native English and 16 native Dutch listeners.
- 3 SNR (0, 8, 16 dB; 6 talker babble)
- 2 syllable positions CV and VC (15 different Vs)

I have (re)analyzed a subset of their data:

- IDs of CV, VC [p], [b], [f], [v], [θ], [ð], [t], [d], [s], [z]
- summed across SNR and vowel, but *not* subject

(see Ashby, Maddox, & Lee, 1994)

Contrast Structure of C Subset



Analysis based on confusions between minimally contrastive pairs, or pairs connected by single lines, e.g., [v]-[z], but *not* [b]-[z]

Confusion Probabilities and Similarity

Confusion probabilities are not readily interpretable.

Similarity Choice Model (Shepard, 1957; Luce, 1963)

$$(1) \qquad p(j \mid i) = \frac{b(j)\eta(i,j)}{\sum_{k=1}^{n} b(k)\eta(i,k)}$$

Symmetric similarity between i and j

$$(2) \qquad \eta(i,j) = \sqrt{\frac{p(i \mid j)p(j \mid i)}{p(i \mid i)p(j \mid j)}}$$

SCM Model Fit

For an n x n confusion matrix, the SCM uses n(n-1)/2 similarity parameters and (n-1) bias parameters. The 'perfect' (or saturated) model has n(n-1) parameters.

So, in the present case, there are 45 similarity parameters and 9 bias parameters.

CV fits ($G^2(36)$) ranged from 13.08 to 36.80 (n.s.)

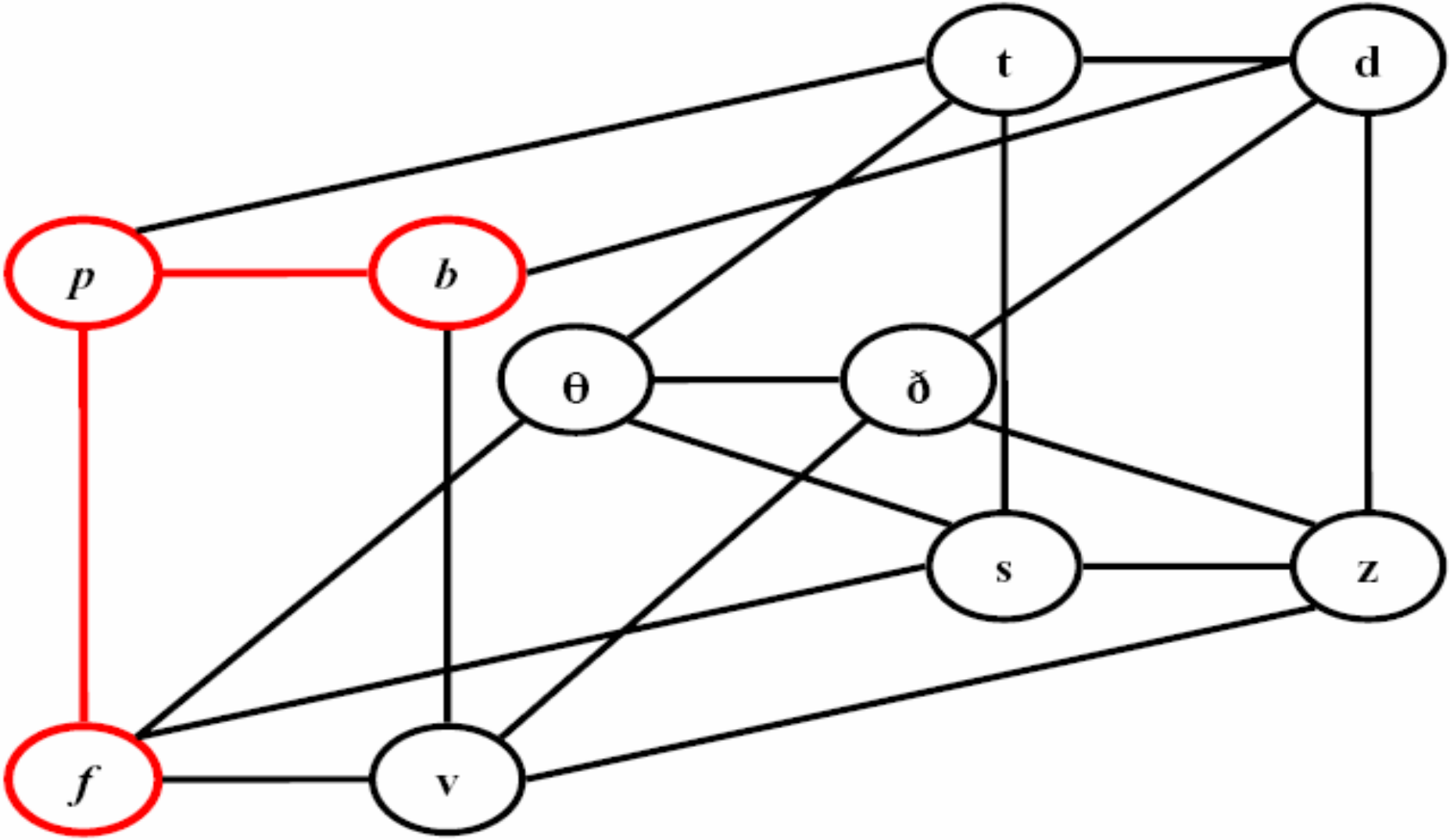
VC fits ($G^2(36)$) ranged from 6.03 to 65.51 (3 $p < 0.05$)

FE, FI, and $\eta(i,j)$

If FE holds, $\eta(i,j)$ with stimuli i and j contrastive along one feature dimension should not differ, on average, from $\eta(i,k)$ with i and k contrastive along a different dimension.

So, to test FE: Compare the 40 pairs of pairs of consonants contrasting on different features and sharing one element, tally the number of the 16 subjects whose data exhibit a given ordinal relationship, e.g., $\eta(\text{p,f}) > \eta(\text{p,b})$. [p]-[f] contrast in manner; [p]-[b] contrast in voicing.

Evaluation of FE & Contrast Space

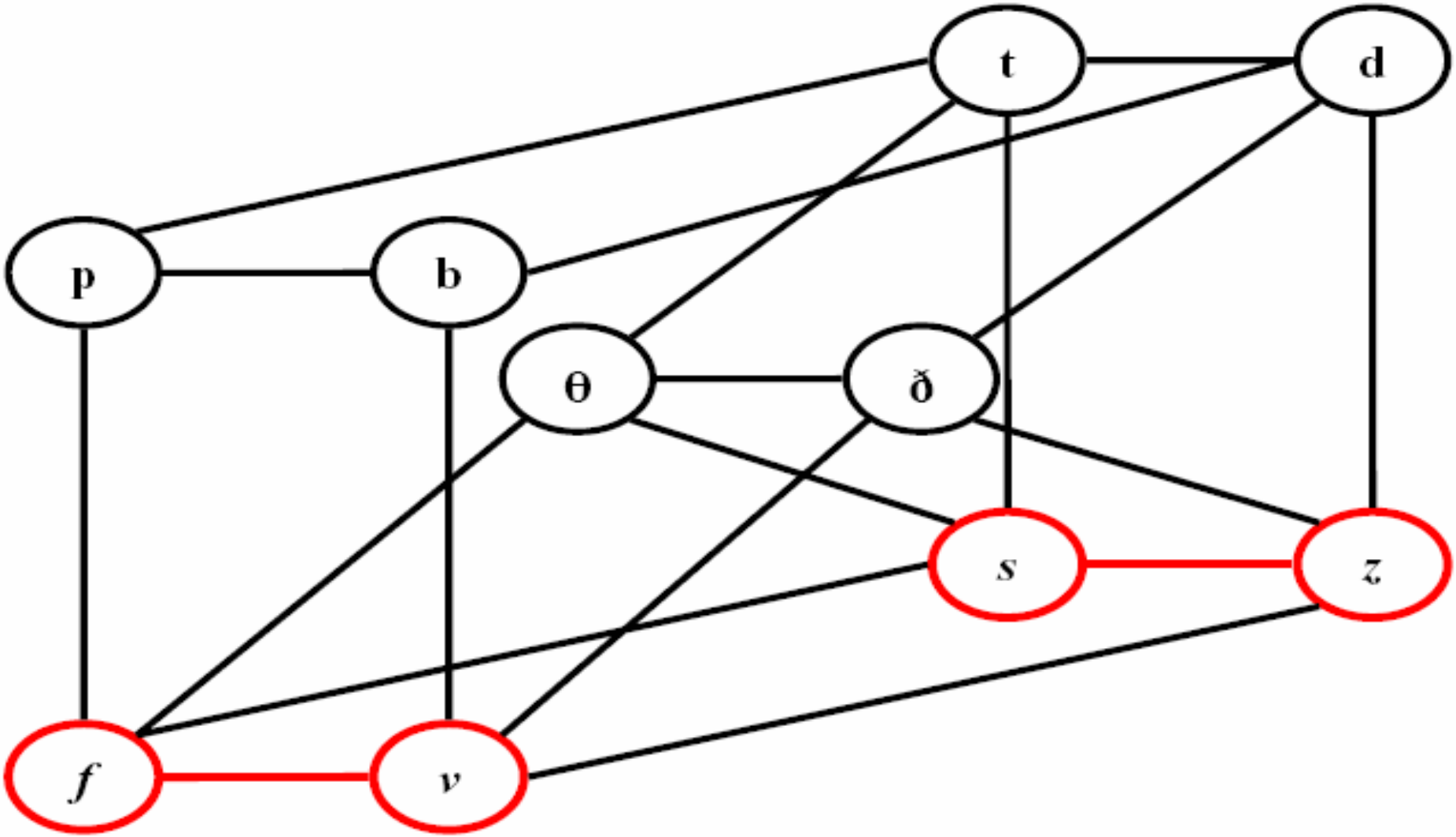


FE, FI, and $\eta(i,j)$

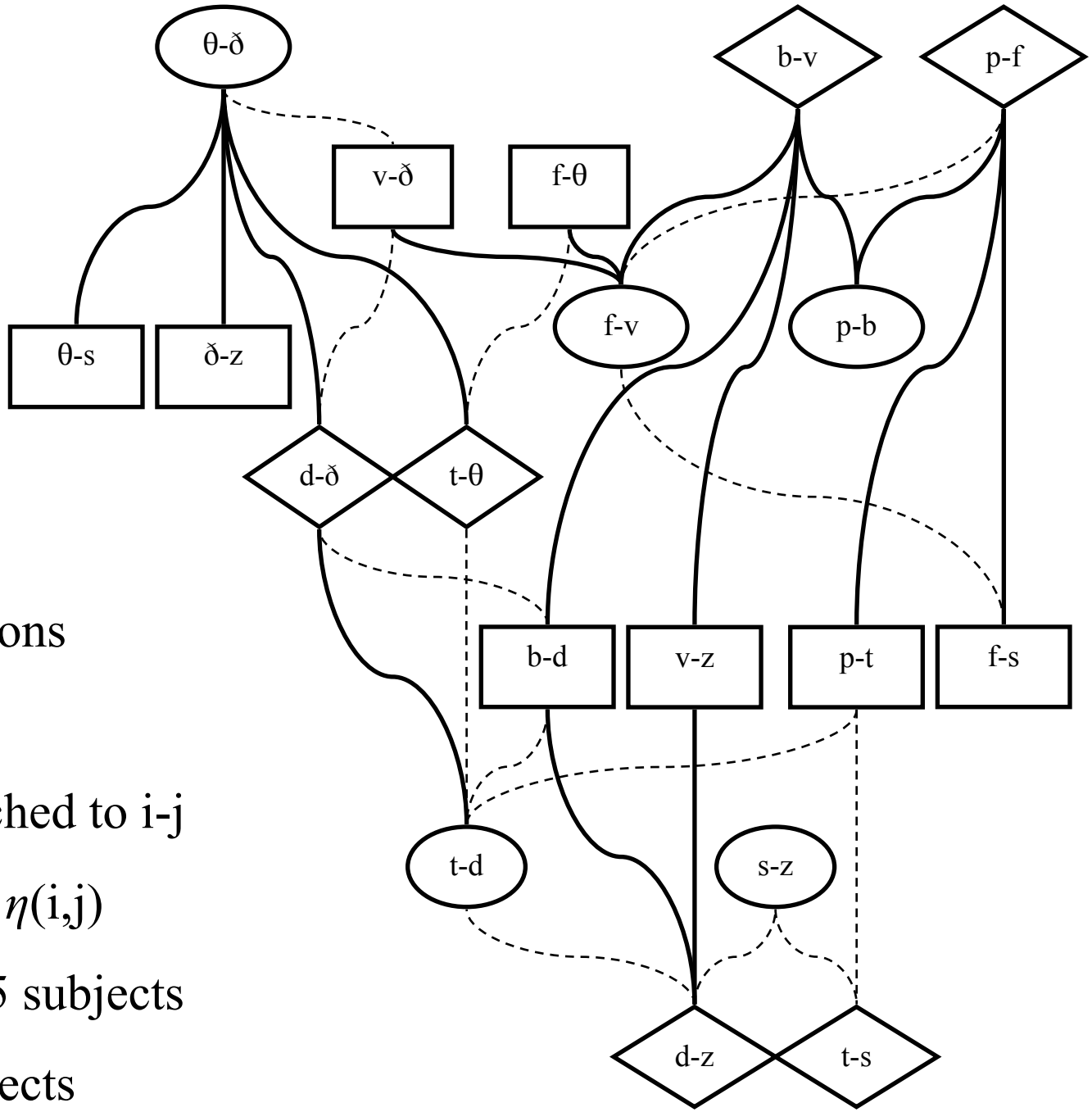
If FI does ***not*** hold, $\eta(i,j)$ with i and j contrastive along one feature dimension should not differ, on average, from $\eta(k,l)$ with k and l contrastive along the same dimension but differing from i and j on a second dimension.

So, to test FI: Compare the 22 pairs of pairs of consonants contrasting along the same dimension and differing on another, tally the number of the 16 subjects whose data exhibit the relationship, e.g., $\eta(\text{f,v}) > \eta(\text{s,z})$. [f]-[v] and [s]-[z] contrast in voicing; they differ in place of articulation.

Contrast Structure of C Subset



Binomial Test Results: Onset



Onset FE comparisons

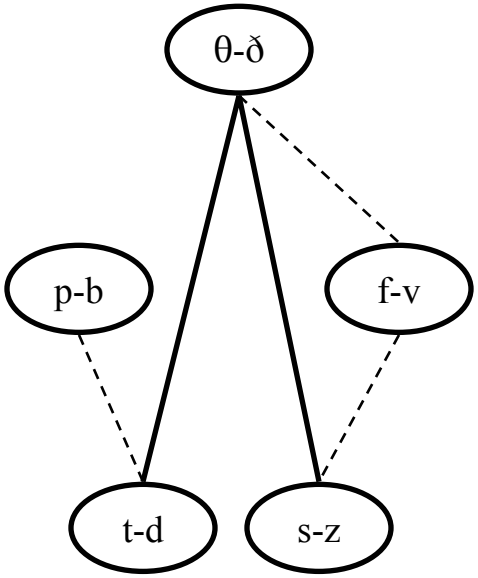
x-y = η(x,y)

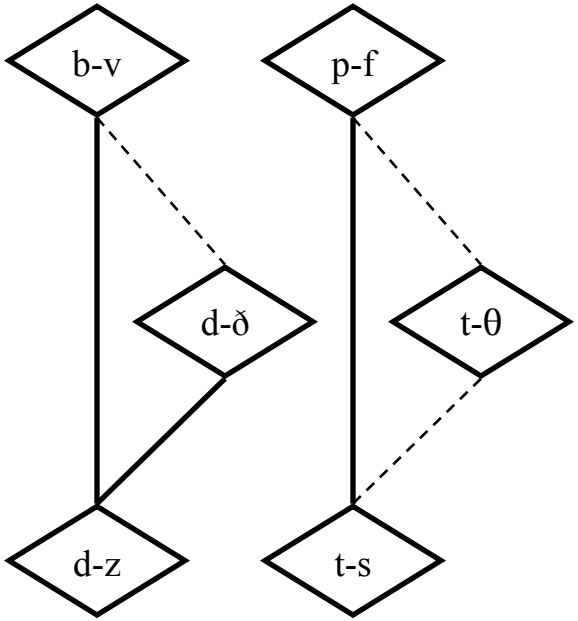
x-y above and attached to i-j

→ η(x,y) > η(i,j)

dashed line = 13-15 subjects

solid line = 16 subjects





Onset FI comparisons

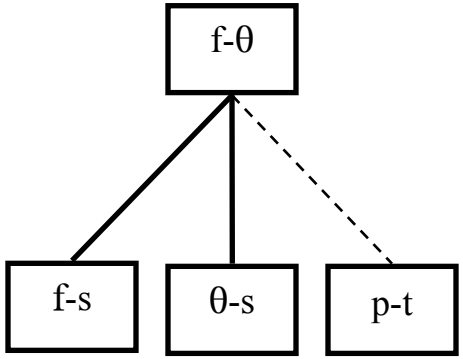
x-y = η(x,y)

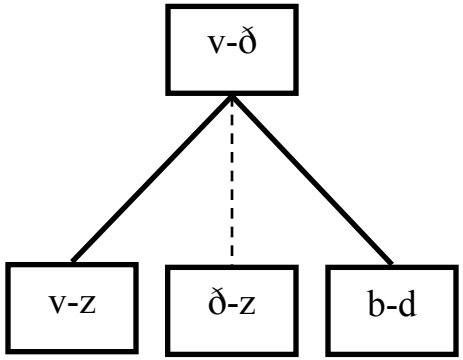
x-y above and attached to i-j

→ η(x,y) > η(i,j)

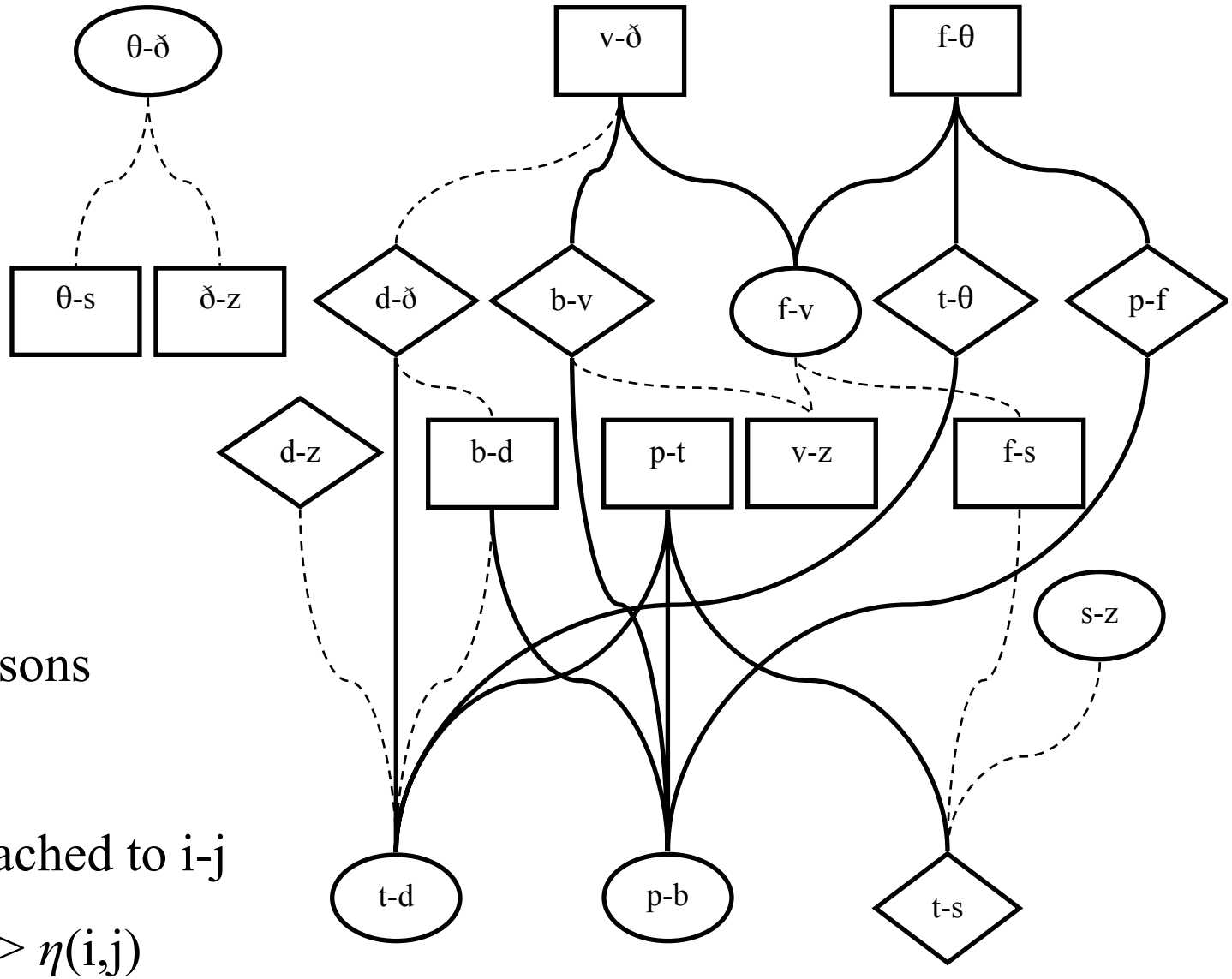
dashed line = 13-15 subjects

solid line = 16 subjects





Binomial Test Results: Coda



Coda FE comparisons

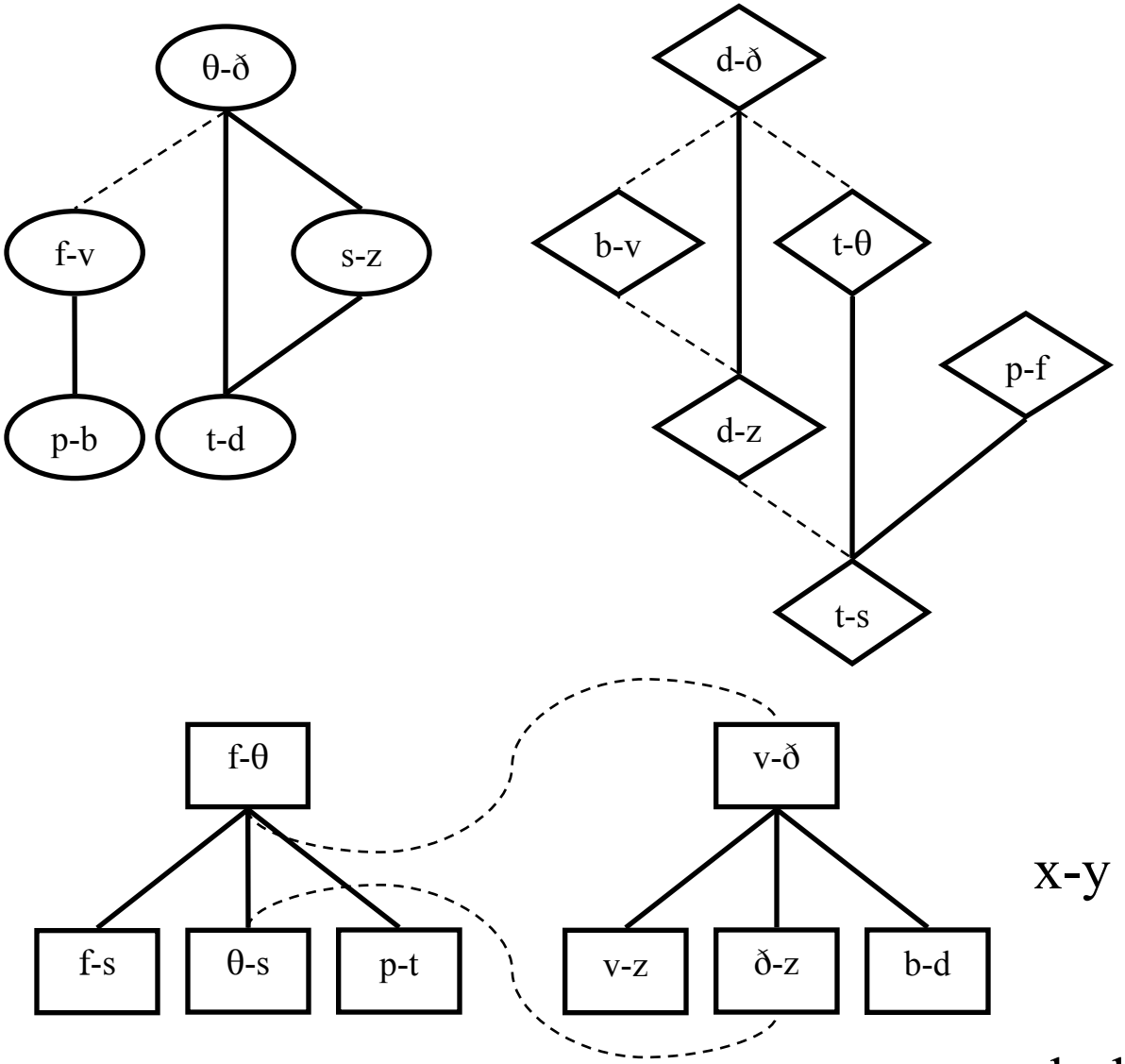
x-y = η(x,y)

x-y above and attached to i-j

→ η(x,y) > η(i,j)

dashed line = 13-15 subjects

solid line = 16 subjects



Coda FI comparisons

x-y = η(x,y)

x-y above and attached to i-j

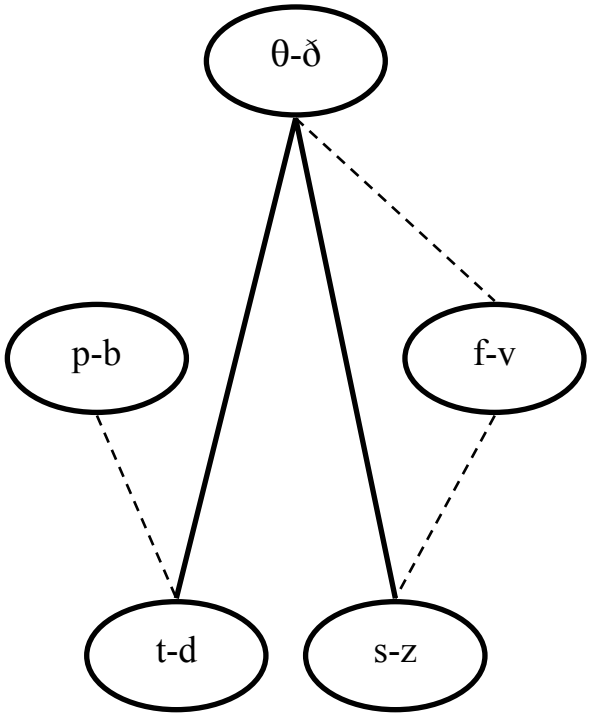
→ η(x,y) > η(i,j)

dashed line = 13-15 subjects

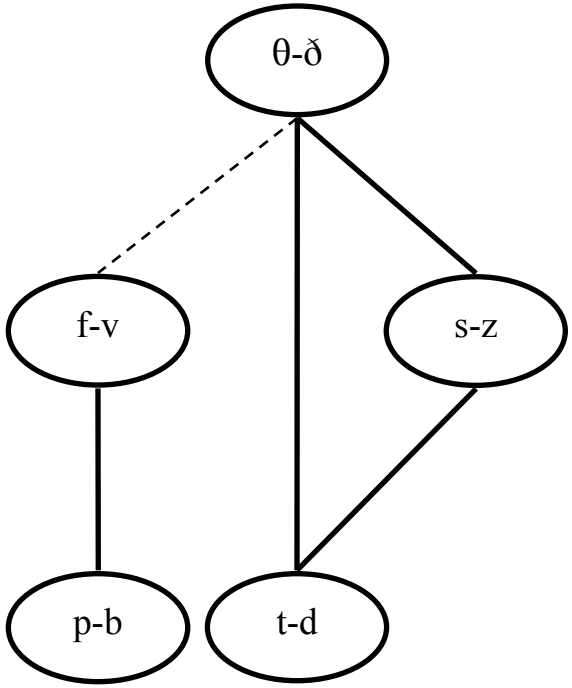
solid line = 16 subjects

FI & Prosodic Context

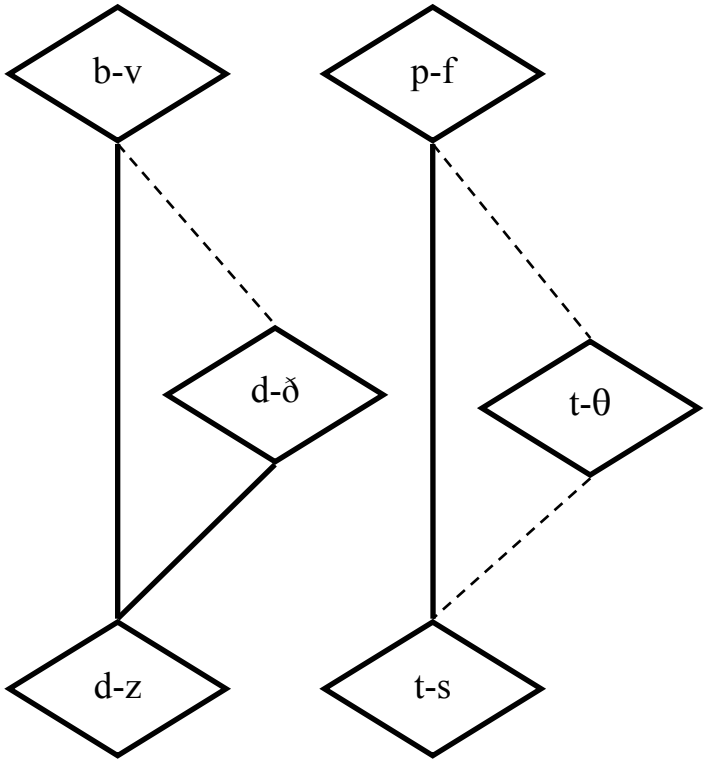
Onset



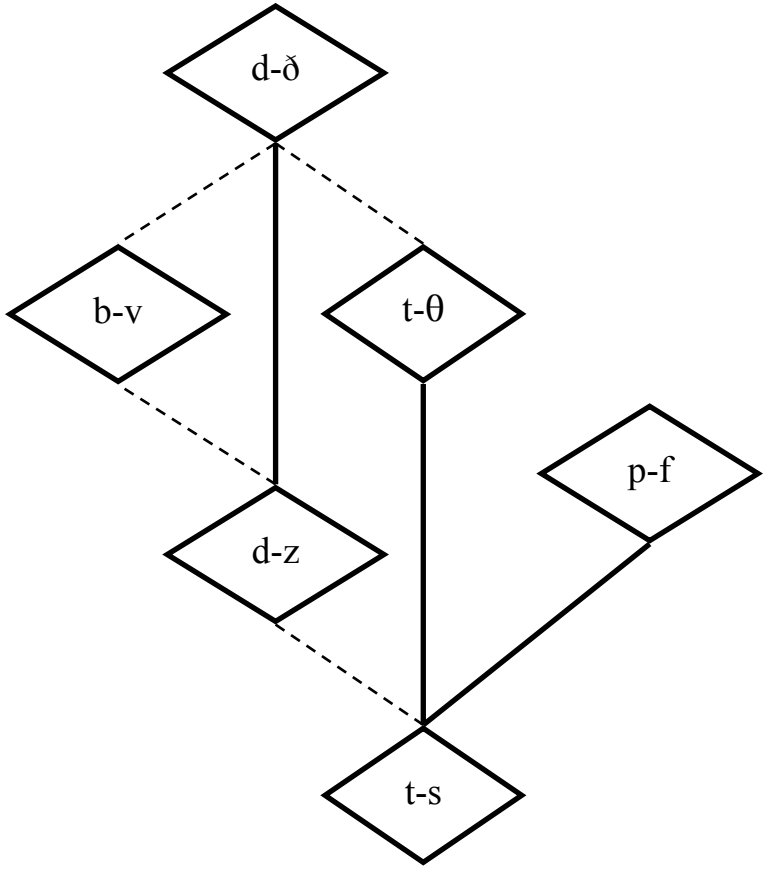
Coda



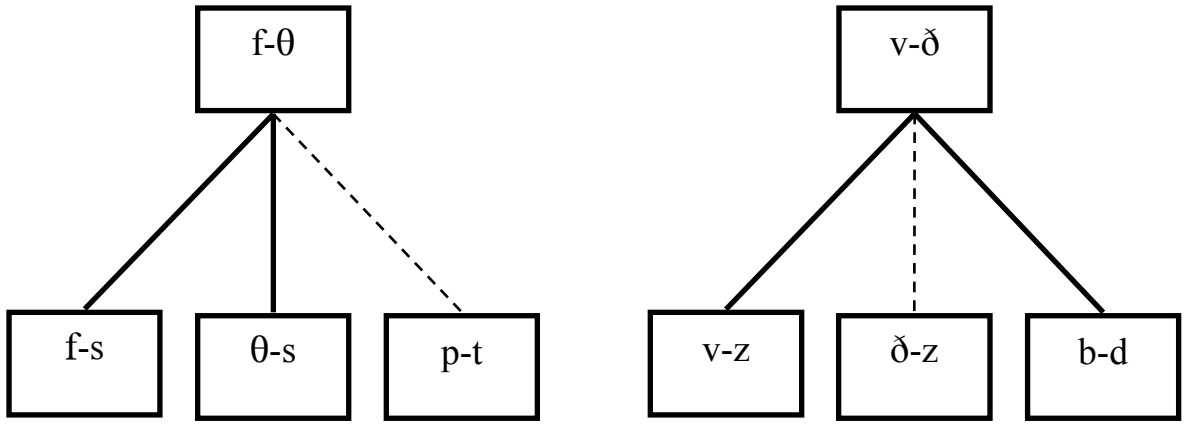
Onset



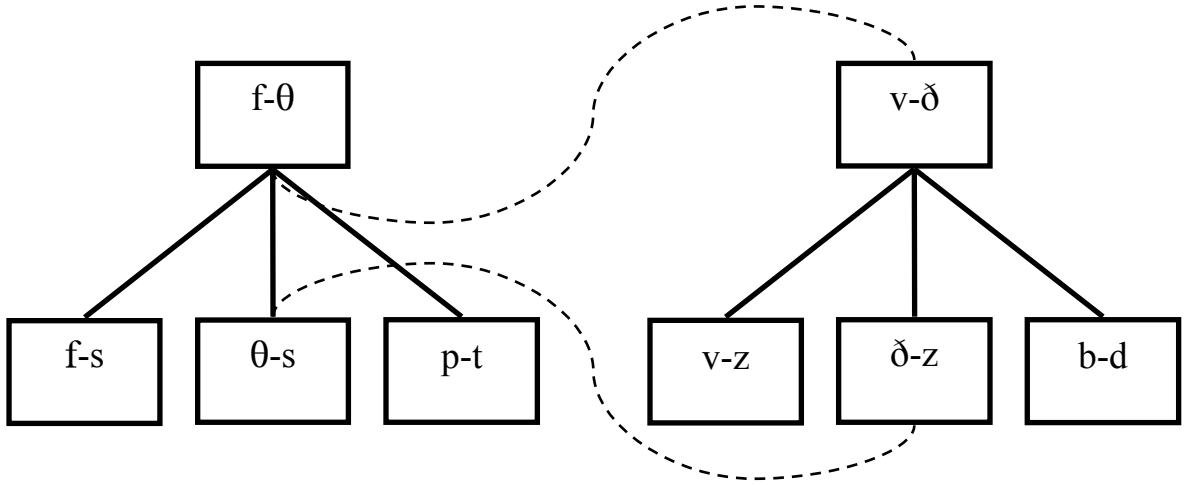
Coda



Onset



Coda



Interim Summary

Tests of FE, FI indicate, indirectly, a role for features through parallel patterns of ordinal similarity relations.

Stronger, more direct evidence that modeling perceptual behavior also captures gross feature-based patterns would bolster our confidence in our findings with regard to FE, FI, and prosodic context.

Individual Differences MDS

Similarity is assumed to be a decreasing function of distance in *m* dimensional space

(3) $$\eta(i,j) = f\left(d(i,j)\right)$$

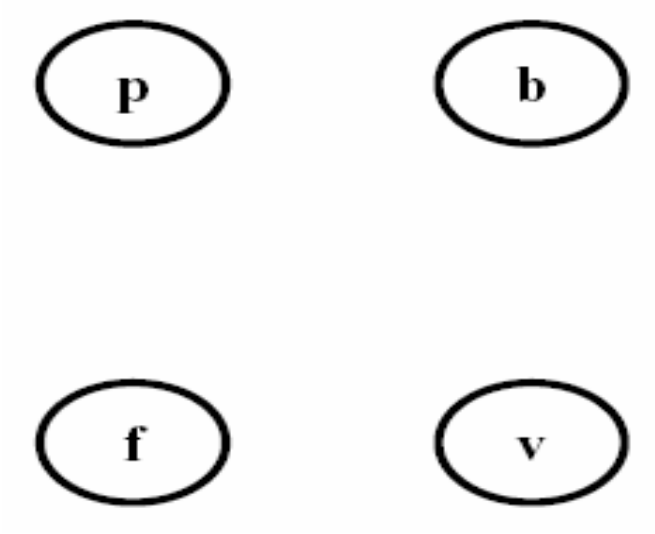
and each dimension may be weighted differently by each subject

(4) $$d_{ijk} = \sqrt{\sum_{m=1}^{M} w_{km}\left(x_{ik} - x_{jk}\right)^2}$$

Individual Differences MDS

Two dimensions – ‘voicing’ and ‘manner’

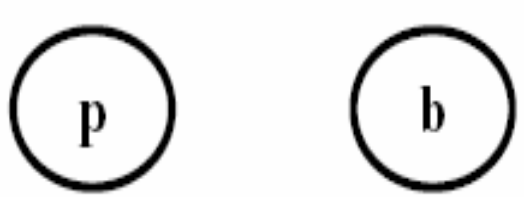
both dimensions are weighted (roughly) equally



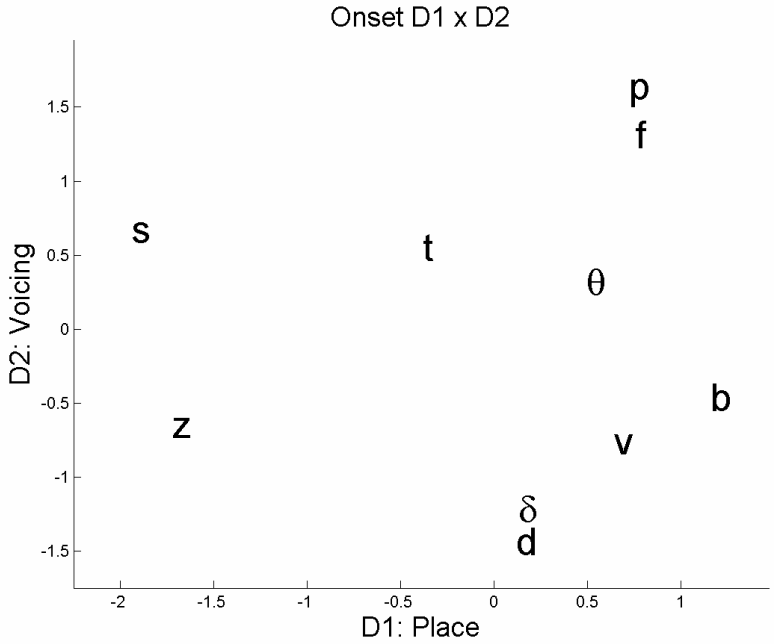
Individual Differences MDS

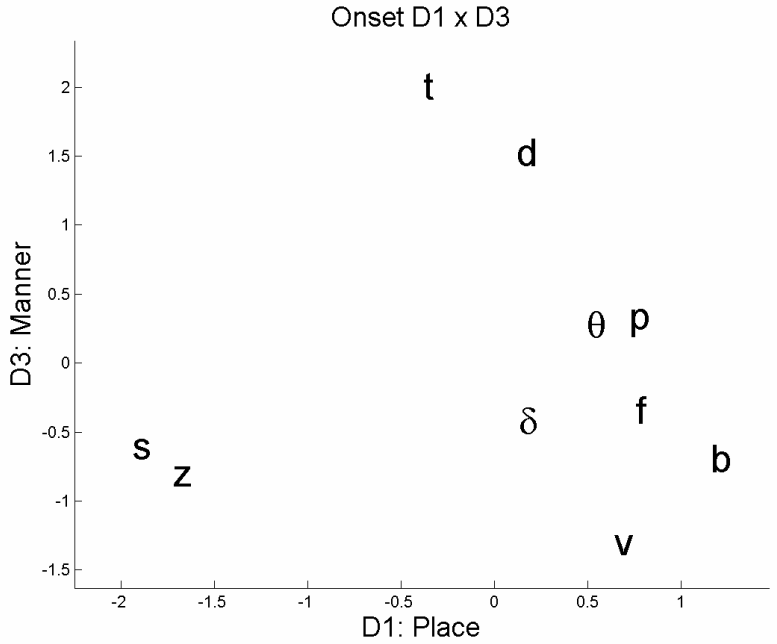
Same two dimensions

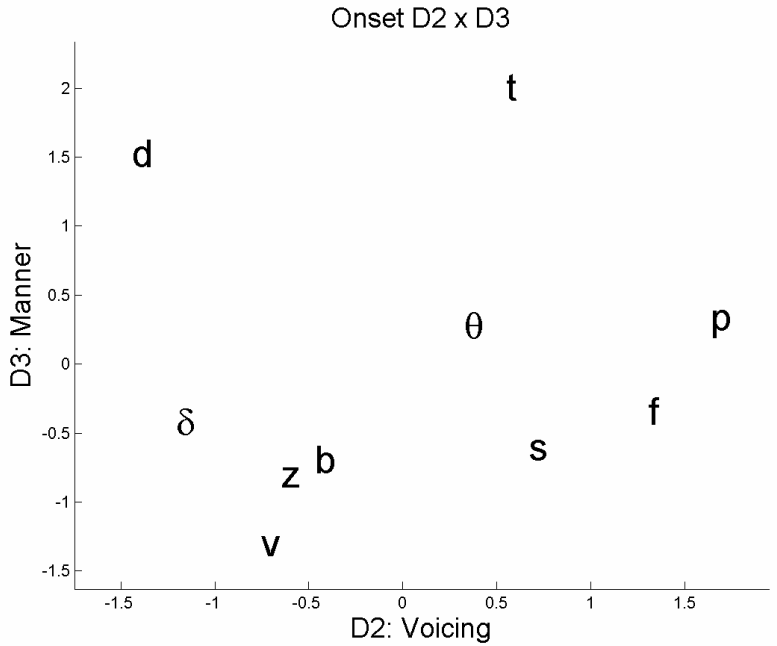
Differential weighting



3D Group Space (INDSCAL), CV

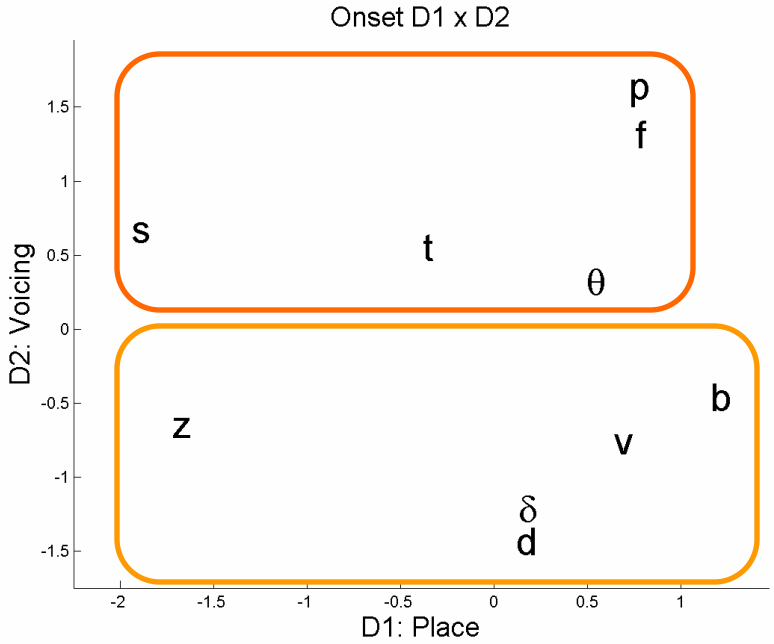


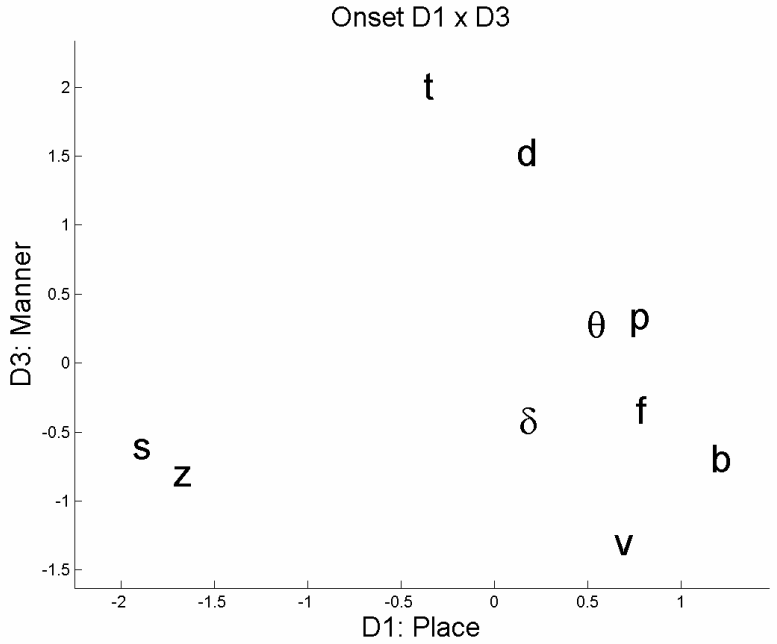


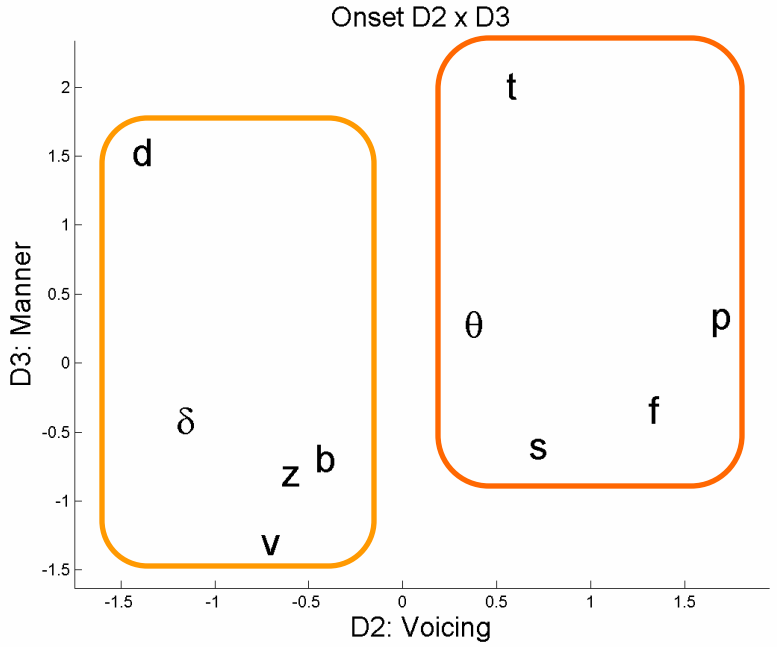


Stress = 0.14739 $R^2 = 0.77032$

3D Group Space (INDSCAL), CV

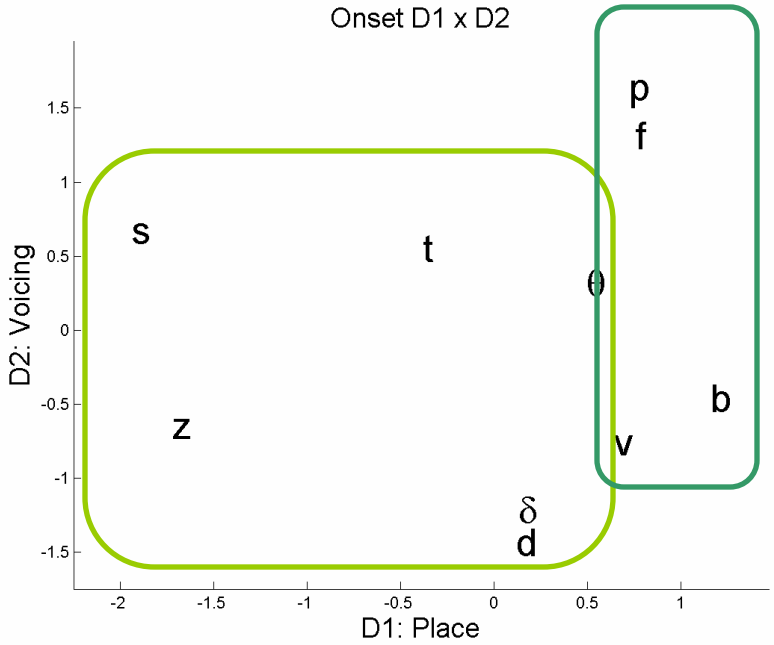


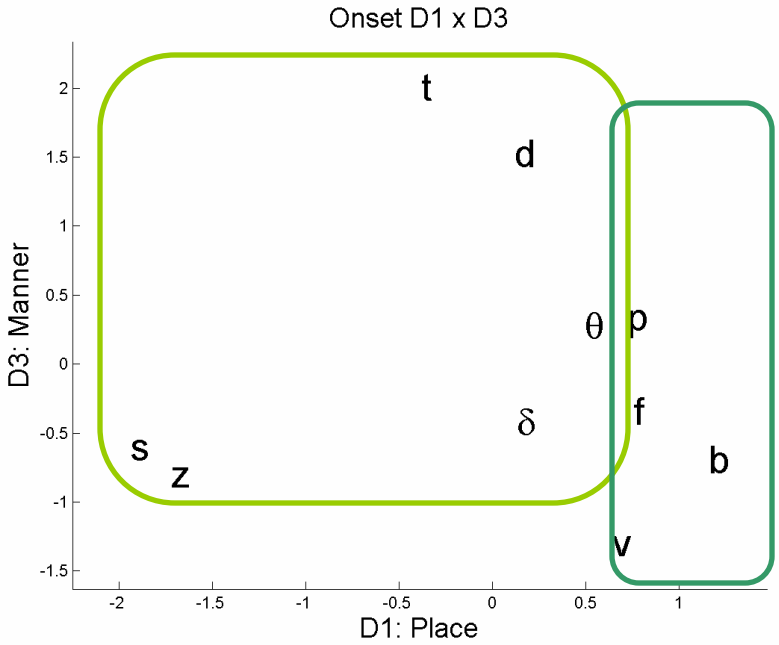


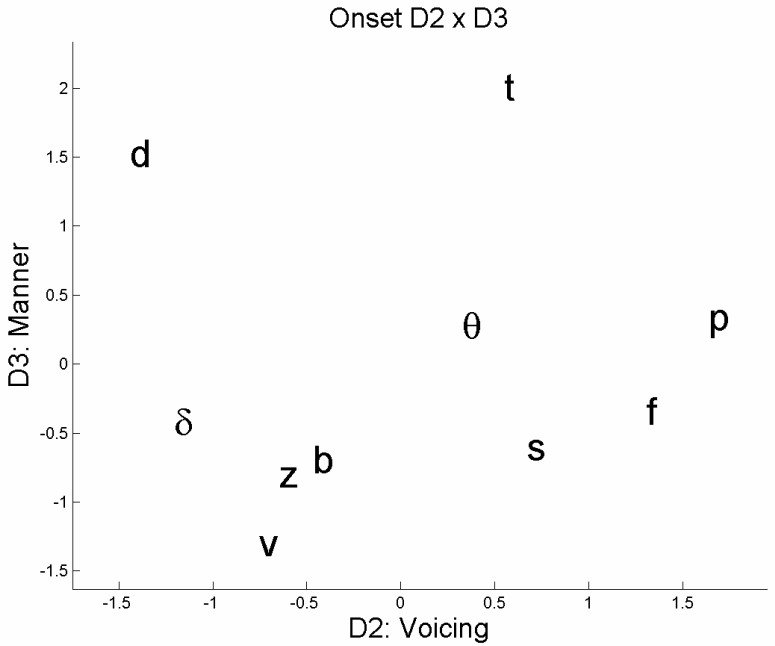


Stress = 0.14739 $R^2 = 0.77032$

3D Group Space (INDSCAL), CV

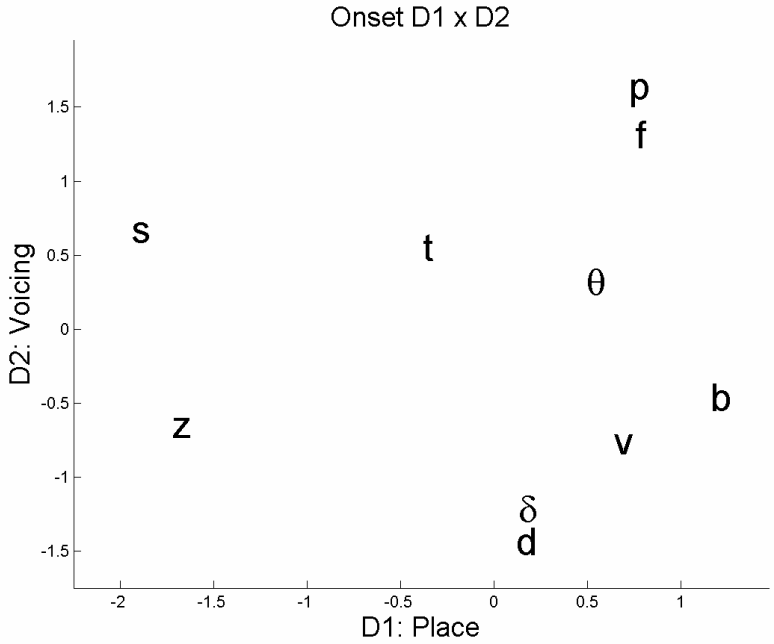


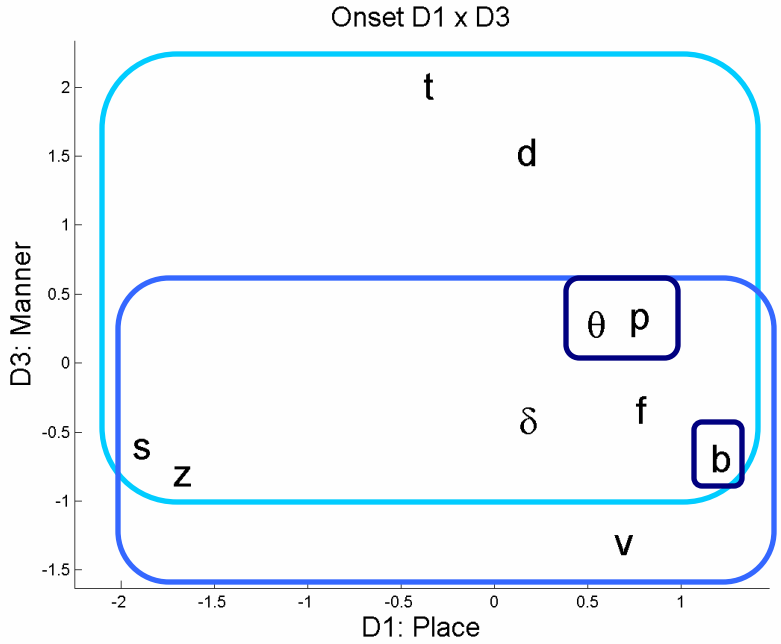


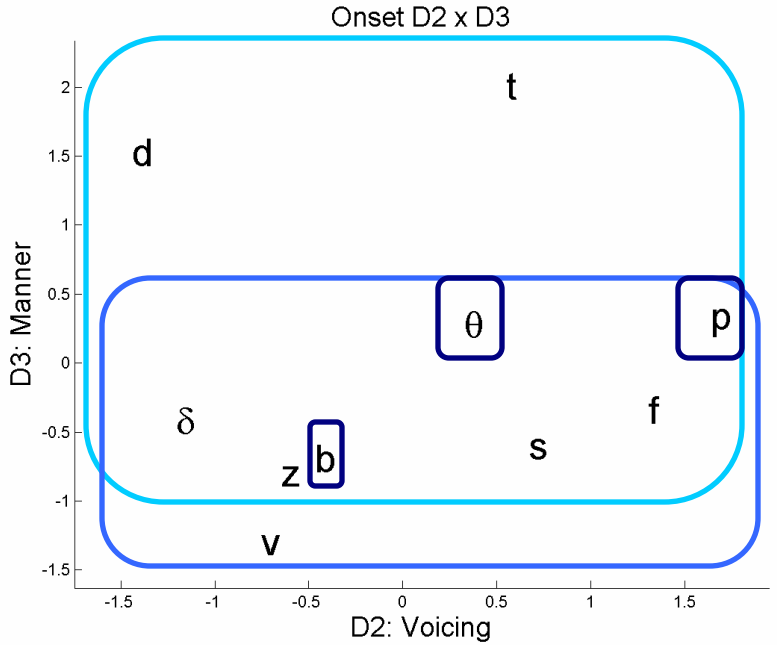


Stress = 0.14739 $R^2 = 0.77032$

3D Group Space (INDSCAL), CV

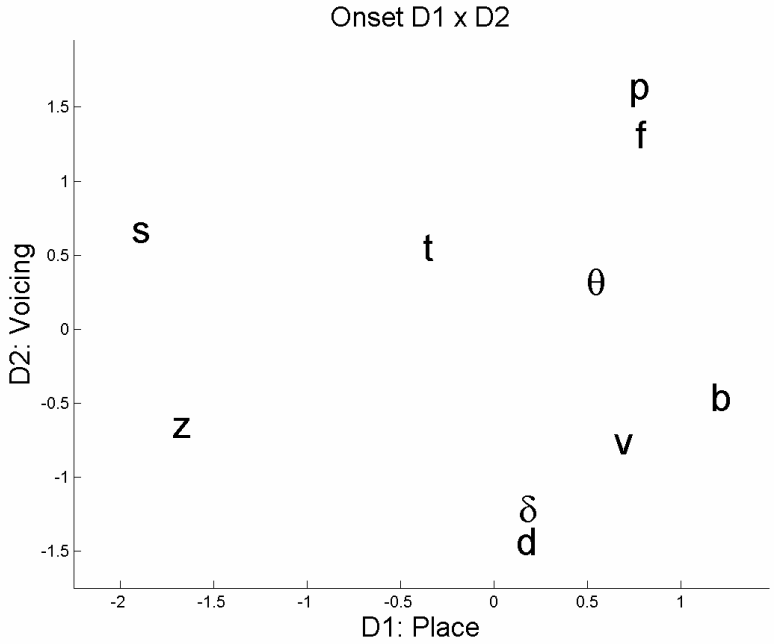


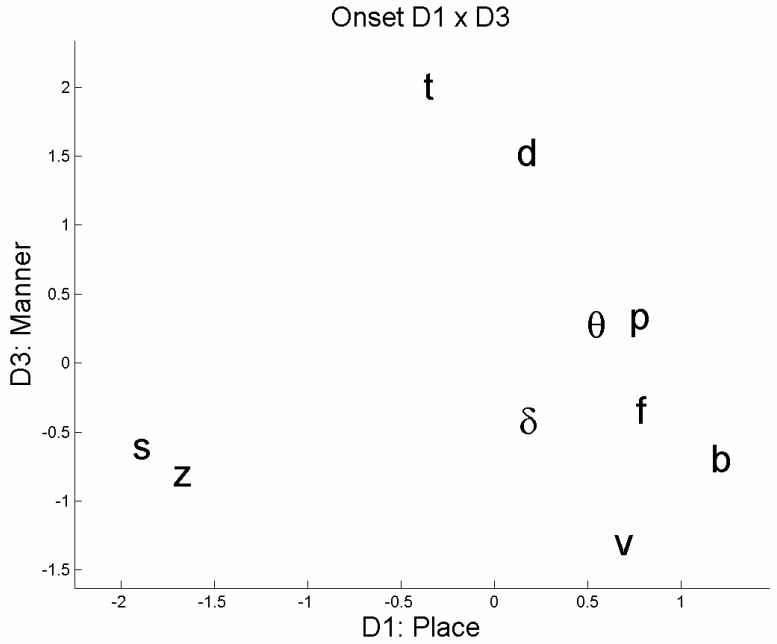


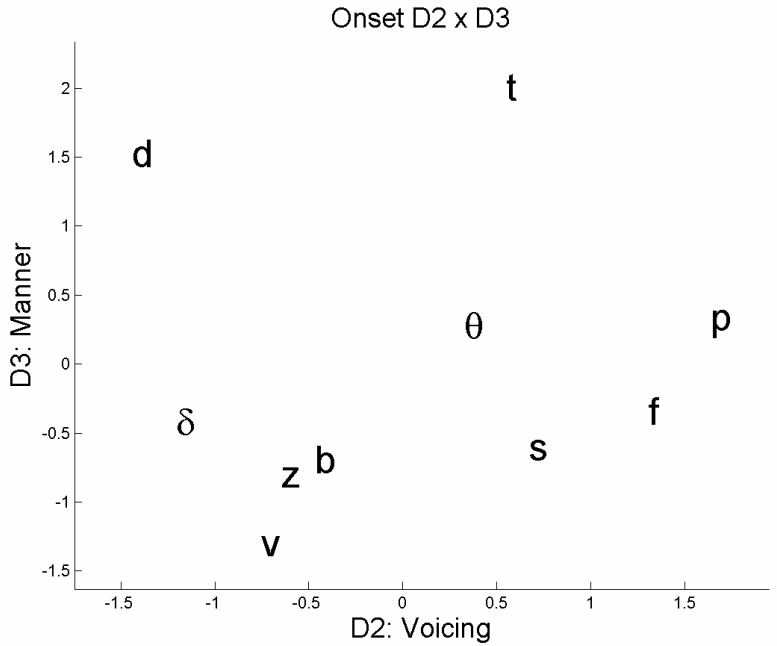


Stress = 0.14739 $R^2 = 0.77032$

3D Group Space (INDSCAL), CV

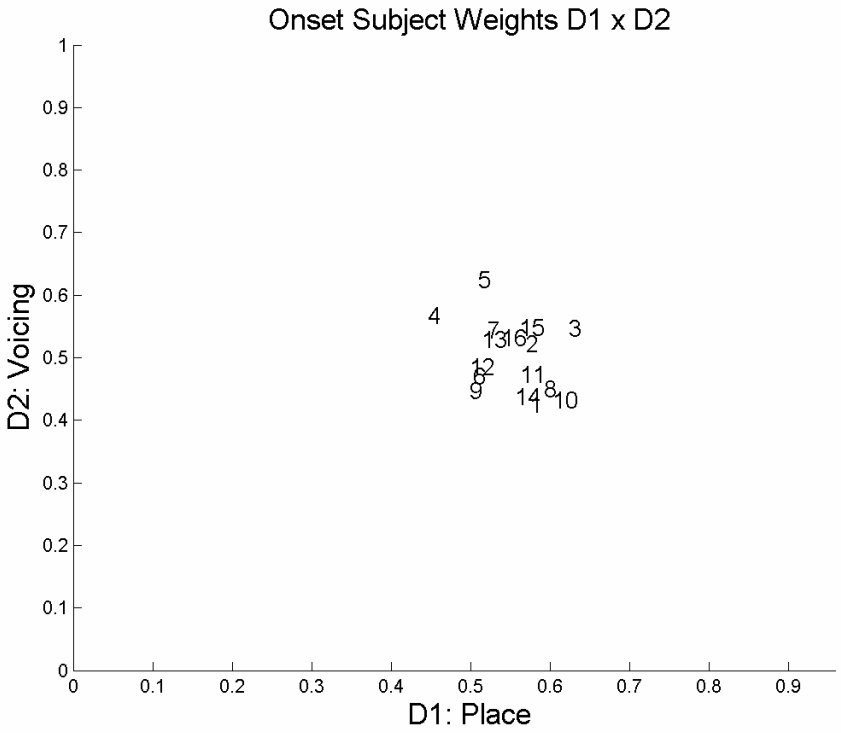


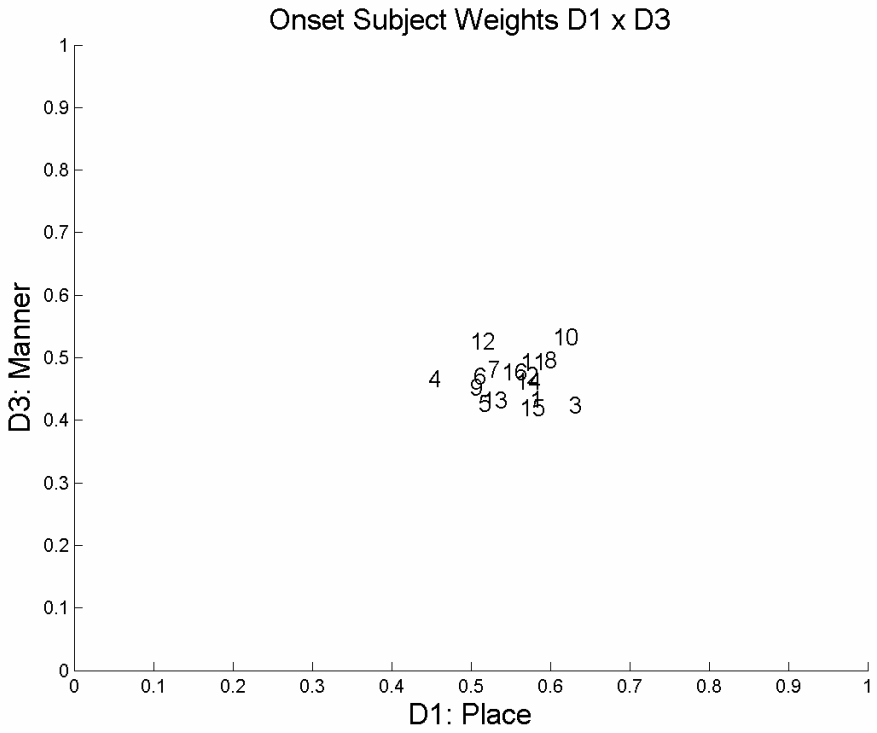




Stress = 0.14739 $R^2 = 0.77032$

Subject Weights, CV

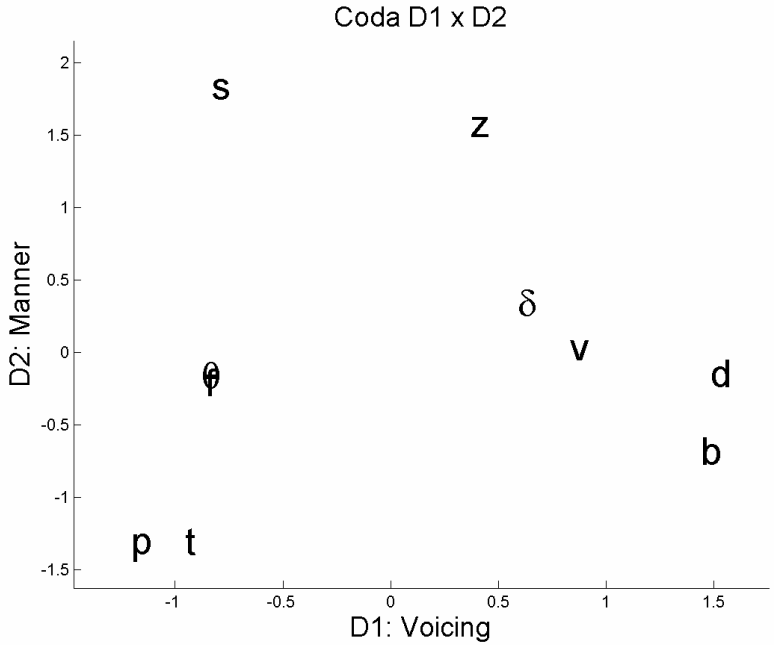


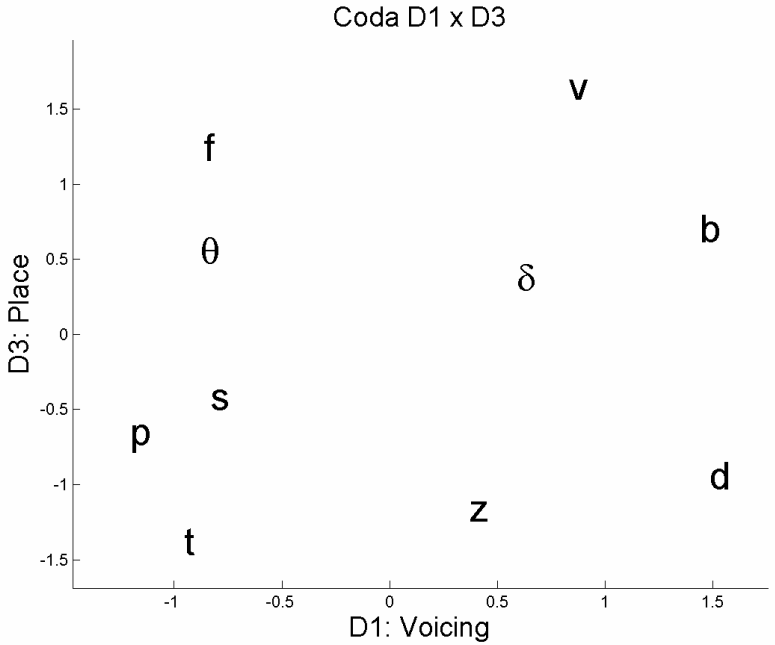


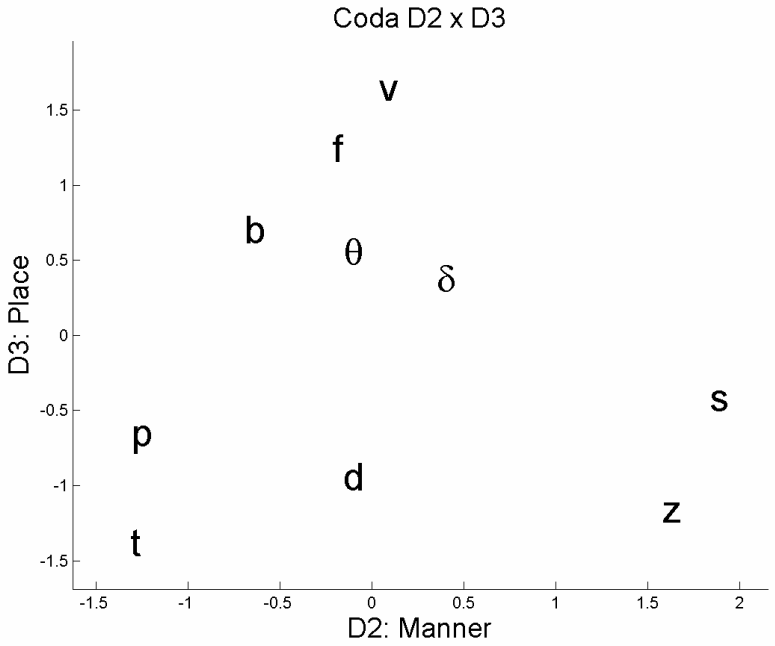
Overall importance of dimensions:

Voicing: 0.2550 Place: 0.2966 Manner: 0.2187

3D Group Space (INDSCAL), VC

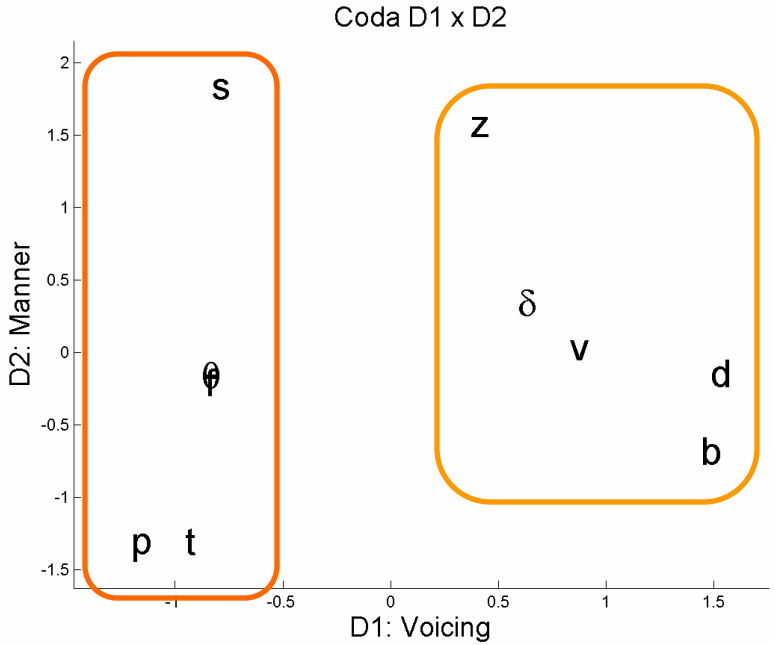


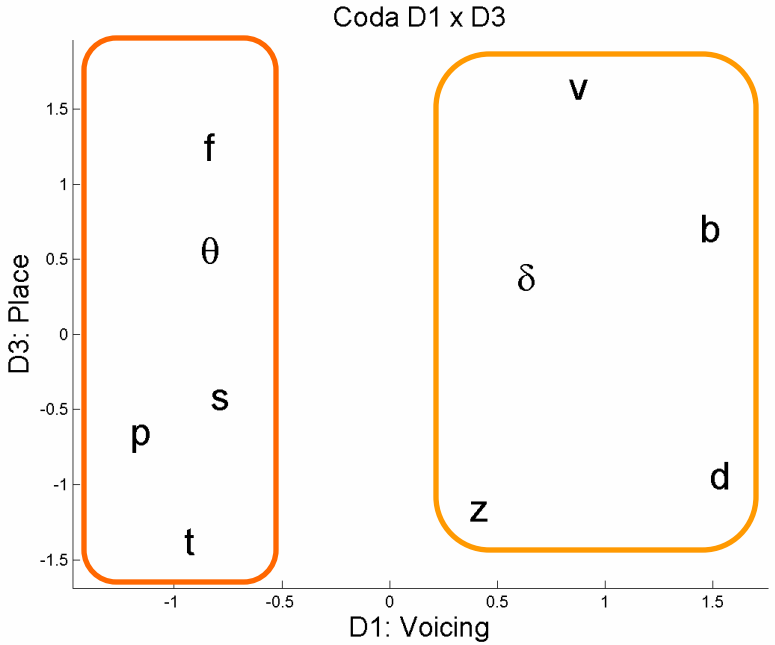


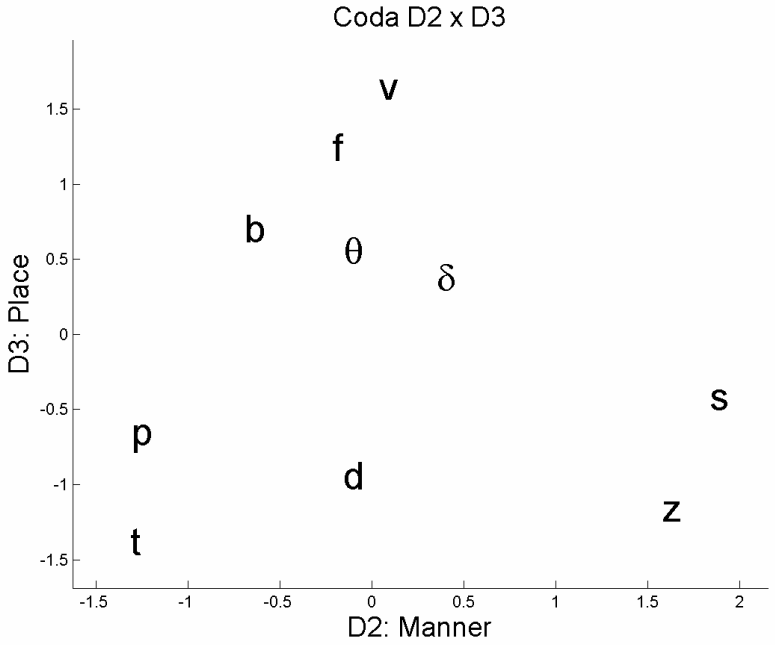


Stress = 0.15058 $R^2 = 0.76859$

3D Group Space (INDSCAL), VC

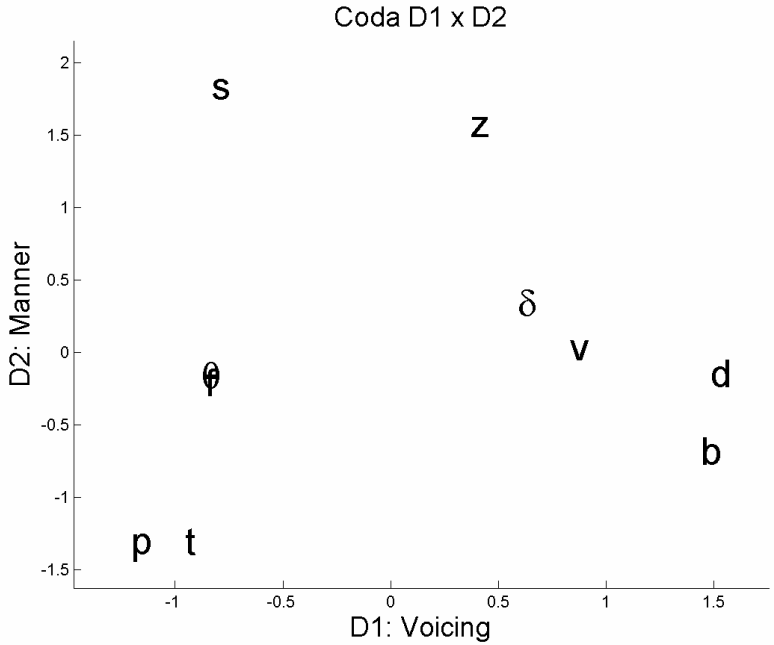


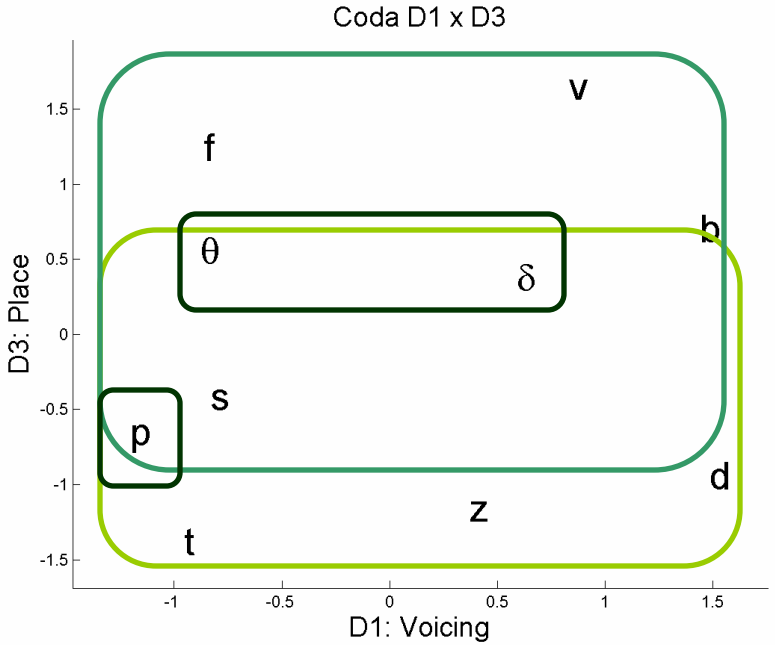


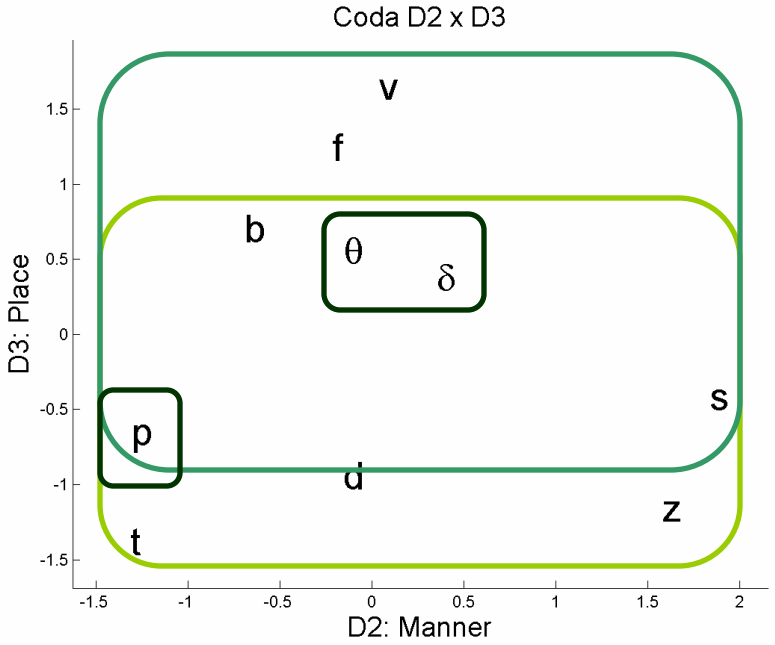


Stress = 0.15058 $R^2 = 0.76859$

3D Group Space (INDSCAL), VC

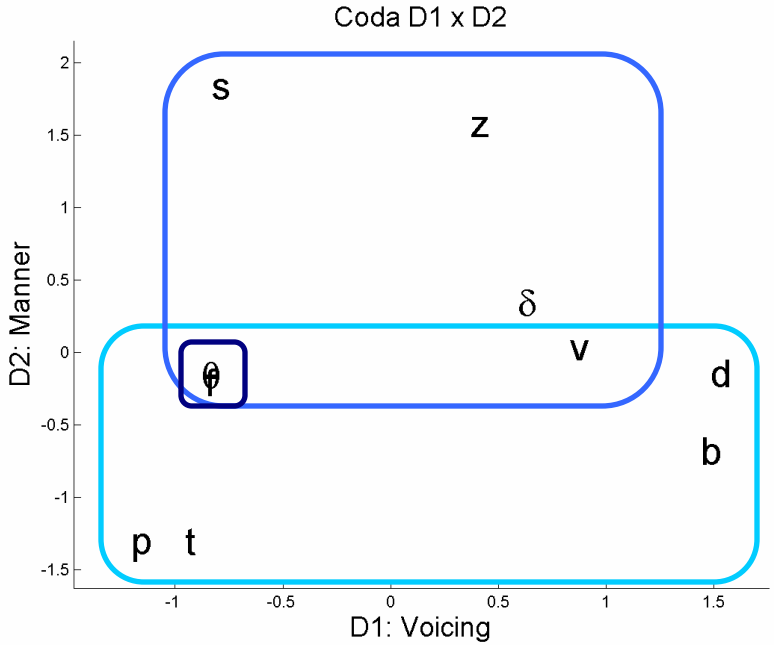


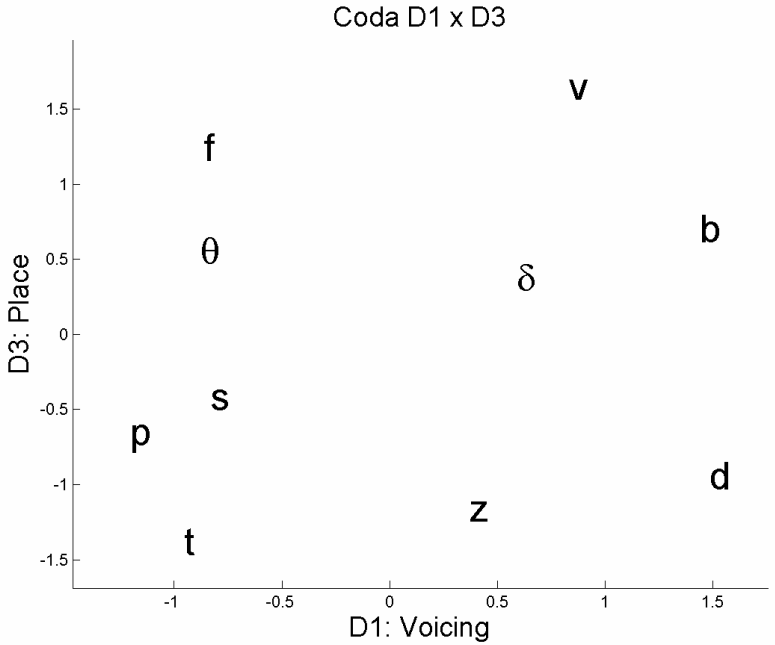


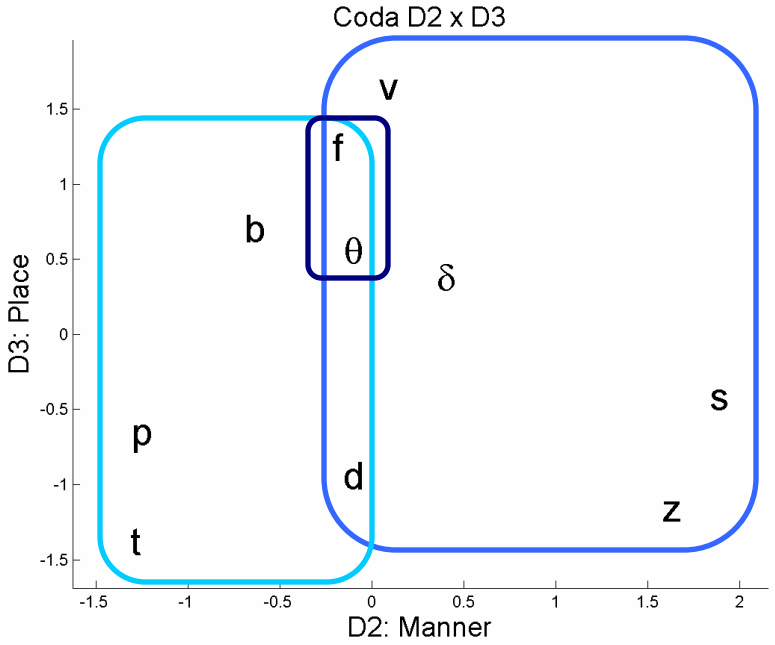


Stress = 0.15058 $R^2 = 0.76859$

3D Group Space (INDSCAL), VC

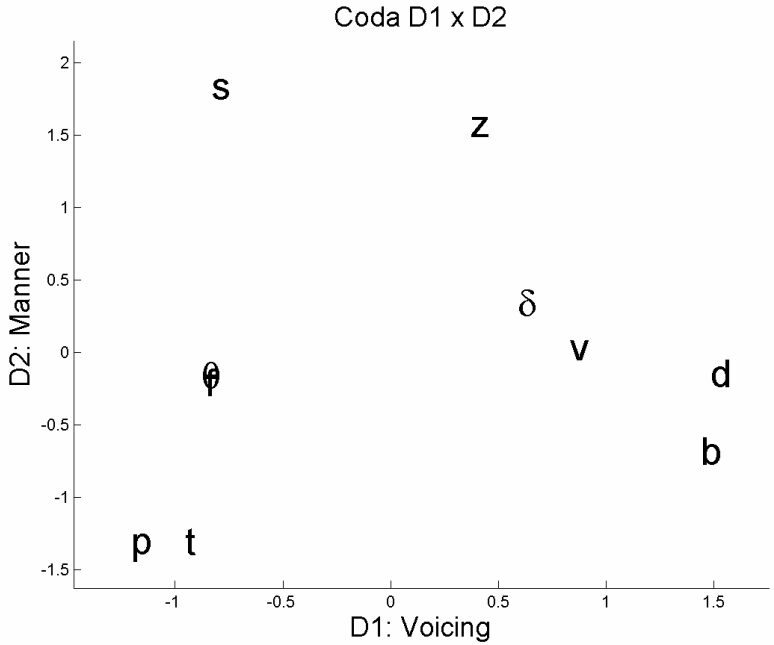


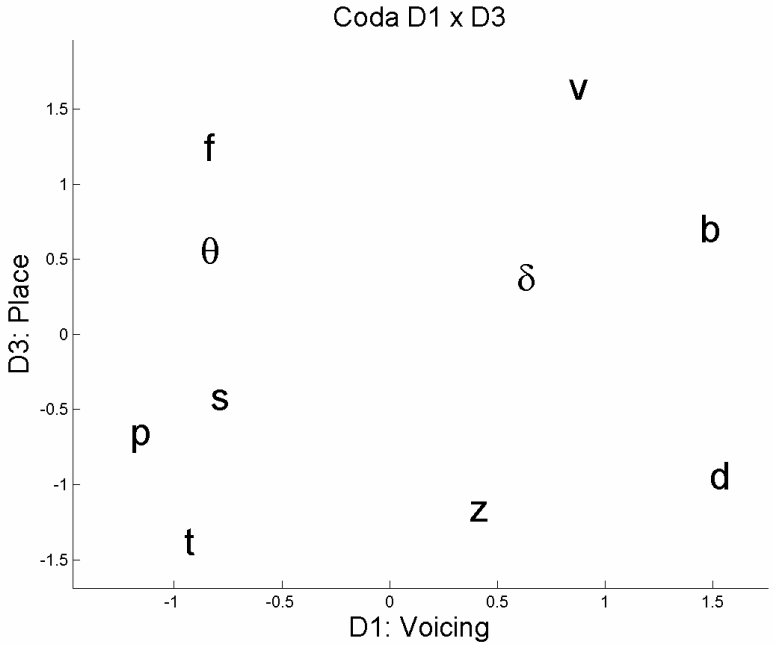


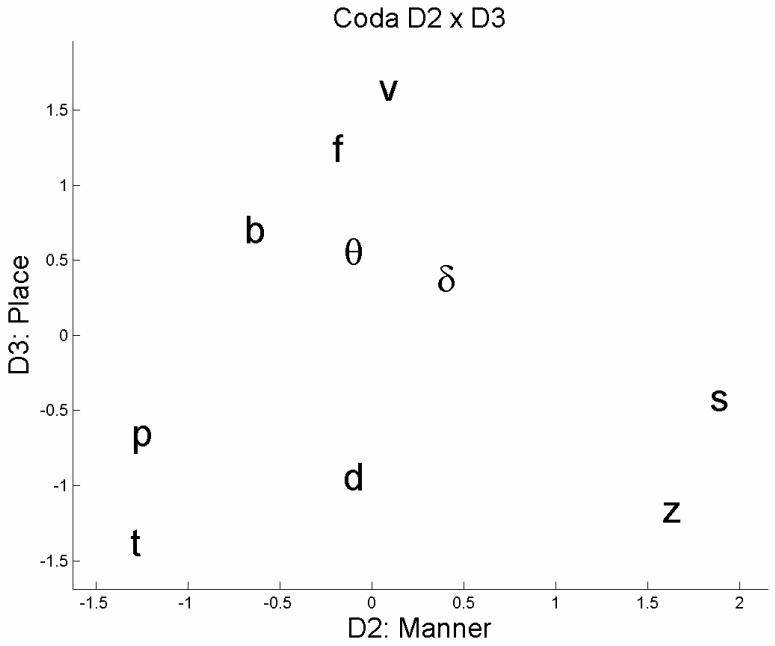


Stress = 0.15058 $R^2 = 0.76859$

3D Group Space (INDSCAL), VC

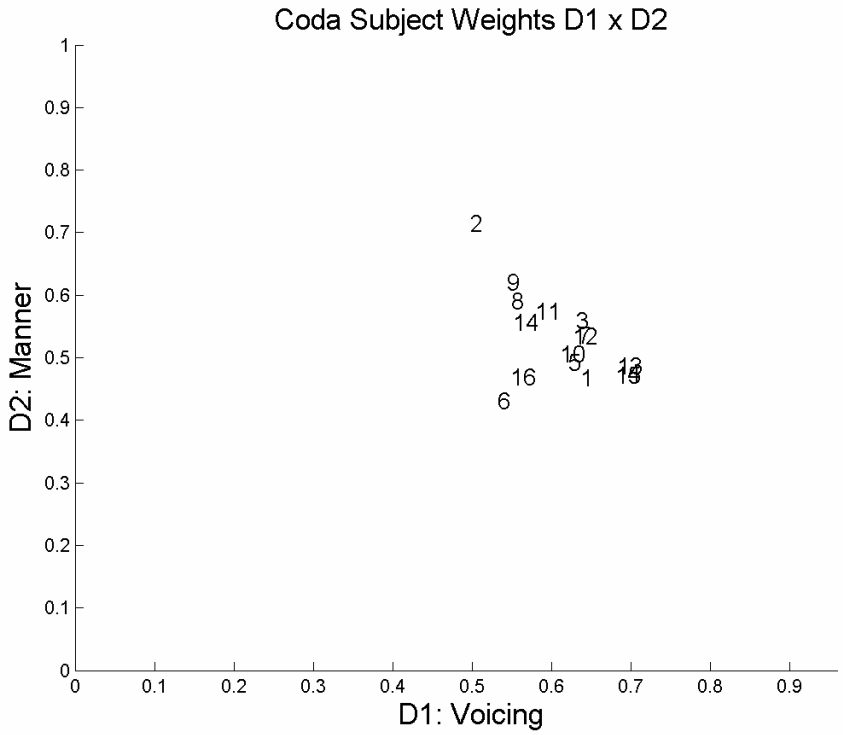


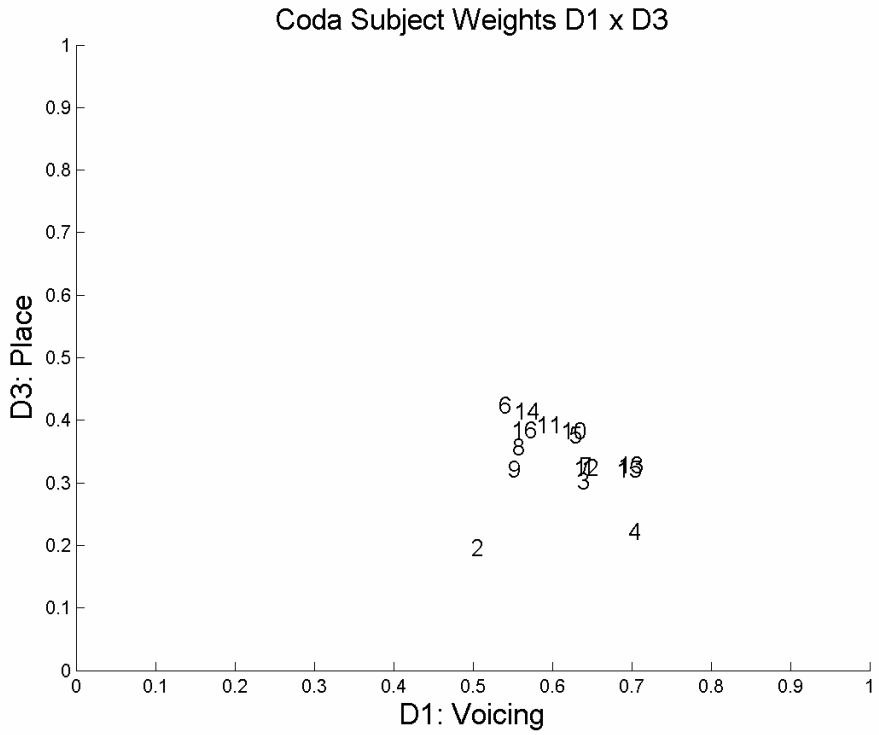




Stress = 0.15058 $R^2 = 0.76859$

Subject Weights, VC





Overall importance of dimensions:

Voicing: 0.3657 Place: 0.1173 Manner: 0.2856

Conclusion and Discussion

Phonological feature theory captures gross categorical structure reasonably well, but misses quite a bit of robust sub-categorical structure in speech perception.

Perceptual modeling captures both.

- the three dimensions of the MDS solutions correspond well to laryngeal, place, and manner features
- interactions and failures of equivalence correspond well to properties of the signal
- the high degree of consistency across subjects indicates that neither the gross nor the fine-grained patterns observed are merely due to idiosyncratic behavior

Conclusion and Discussion

The observed feature interactions make segments look like something more than simple lists (or hierarchical ‘trees’) of features.

Systematic variation in inter-segment similarity has implications for models of word recognition (e.g., the Neighborhood Activation Model) and language change

- some neighbors are ‘closer’ in than others
- neighborhood structure depends, in part, on prosodic context (onset vs. coda)
- large differences in perceptual similarity across contrasts may correspond to differences in patterns of historical contrast loss

Conclusion and Discussion

Future research directions

- consonant identification confusion data collected explicitly for the purpose of perceptual modeling would enable modeling of (differences in) similarity magnitudes
- various perceptual models provide methods for testing specific hypotheses regarding perceptual independence and perceptual separability of features, decision factors
- SCM analyses like those employed here can be extended to other contrasts (e.g., +/- nasal, place)
- SCM bias parameters can be analyzed with regard to, e.g., segment frequency

A Parting Question

Do language learners start with features and learn to map them onto produced and perceived speech, or do they start with, and abstract away from, speech to build the categories?

Although the present data do not answer this question, they do underscore the importance of it.

References

Ashby, F. G., Maddox, W. T., & Lee, W. W. (1994). On the dangers of averaging across subjects when using multidimensional scaling or the similarity-choice model. *Psychological Science*, 5(3), 144-151.

Cutler, A., Weber, A., Smits, R., & Cooper, N. (2004). Patterns of English phoneme confusions by native and non-native listeners. *JASA*, 116(6), 3668-3678.

Goldstein, L. (1980). Categorical features in speech perception and production. *JASA*, 67(4), 1336-1348.

Klatt, D. H. (1968). Structure of confusions in short-term memory between English consonants. *JASA*, 44(2), 401-407.

Luce, R. D. (1963). Detection and recognition. In R. D. Luce, R. R. Bush, & E. Galanter (Eds.), *Handbook of Mathematical Psychology, v.1* (pp. 103-189), New York: Wiley.

Miller, G. A., & Nicely, P. E. (1955). An analysis of perceptual confusions among some English consonants. *JASA*, 27(2), 338-352.

Shepard, R. N. (1957). Stimulus and response generalization: A stochastic model relating generalization to distance in psychological space. *Psychometrika*, 22(4), 325-345.

Shepard, R. N. (1972). Psychological representation of speech sounds. In E. E. David, Jr., & P. B. Denes (Eds.), *Human Communication: A unified view* (pp. 67-113), New York: McGraw-Hill.

Shepard, R. N., & Arabie, P. (1979). Additive clustering: Representation of similarities as combinations of discrete overlapping properties. *Psychological Review*, 86(2), 87-123.

Soli, S. D., & Arabie, P. (1979). Auditory versus phonetic accounts of observed confusions betweeen consonant phonemes. *JASA*, 66(1), 46-59.

Soli, S. D., Arabie, P., & Carroll, J. D. (1986). Discrete representation of perceptual structure underlying consonant confusions. *JASA*, 79(3), 826-837.

Wang, M. D., & Bilger, R. C. (1973). Consonant confusions in noise: A study of perceptual features. *JASA*, 54(5), 1248-1266.

Hierarchical Model Fitting

Analysis of response bias via rank order correlation between segment type frequency, segment token frequency, and token-frequency-weighted type frequency should shed light on any relationship that exists.

Hierarchical model fitting – encoding feature (mis)match and feature interactions directly

$$\eta(i,j) = \prod_{n} f_n$$