2aSC27. Statistical Relationships in Distinctive Feature Models and Acoustic-Phonetic Properties of English Consonants

Kenneth J. de Jong
Kirsten T. Regier, & Aaron Albin
Department of Linguistics, Indiana University

Noah H. Silbert
Center for the Advanced Study of Language
University of Maryland

Main Points

* This study examines the extent to which more general, and less general representations of consonants can account for a corpus of acoustic measures of consonants
* Fitted models which rely on generalized representations perform more poorly than more specific models
* Patterns suggest that generalized representations of the consonants are not transparently available in their acoustics, but need to be constructed either by individual cognitive systems or by collective language communities.

Introduction

Traditional Consonant Representations

* SEGMENTAL (Alphabetic) model:
Consonants are segments chosen from inventory

/ b, p, d, t, g, k, v, f, z, s, ʒ, ʃ, dʒ, tʃ, m, n …/

* FEATURAL (Feature bundle) model:
Consonants are collocations of binary feature values

	b	p	d	t	g	k	v	f	z	s	ʒ	ʃ	dʒ	tʃ	m	n
Voicing	+	-	+	-	+	-	+	-	+	-	+	-	+	-	+	+
Coronal	-	-	+	+	-	-	-	-	+	+	+	+	+	+	-	+
Anterior	+	+	+	+	-	-	+	+	+	+	-	-	-	-	+	+
Sonorant	-	-	-	-	-	-	-	-	-	-	-	-	-	-	+	+
Continuant	-	-	-	-	-	-	+	+	+	+	+	+	-	-	-	-
Strident	-	-	-	-	-	-	+	+	+	+	+	+	+	+	-	-

* ORDINAL (Phonetic chart) model:
Consonants are addresses in a 3- to 4- dimensional matrix

	Labial		Alveolar		Palato-Alveolar		Velar	
	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless	Voiced	Voiceless
Plosive	b	p	d	t			g	k
Fricative	v	f	z	s	ʒ	ʃ		
Affricate					dʒ	tʃ		
Nasal	m		n					

Relations between Representations

* Trade-off between *specific* (providing much flexibility for individual idiosyncrasy), and *general* (providing much more compact representation for the system)
* Most specific: SEGMENTAL:
Each consonant requires its own specification
* Most general: ORDINAL:
Each requires only (in example above) 3 specifications

General Question

To what extent are the properties of these representations evident in aspects of human behavior? Previous work: Perceptual confusions

* Silbert & de Jong (2007, to appear): General feature structure is evident in similarity parameters from segment identification data taken from Cutler et al (2004)
* Silbert (in review a & b): many variations from featural structure are robustly apparent in confusion data

Current Question

To what extent are the properties of these representations evident in production acoustics?

* There are many relevant acoustic dimensions in the speech signal; distinctive feature structure is also multi-dimensional; the mapping between the two is many-to-many.
* With respect to **Best Estimates of Segmental Categories**, are consonant categories arranged in a multi-dimensional acoustic space according to the more general structures (FEATURAL or ORDINAL)?

* With respect to **Category Internal Variance**, is the internal structure of the consonant categories shared across categories? Or does each consonant have its own distribution in the multi-dimensional space?

Methods

Corpus

* 16 English consonants, as listed to the left
* Preceding (ONSET) and following (CODA) vowel /a/
* Produced in isolation
* 10 repetitions of each target pseudo-randomized
* 3200 tokens per prosodic position, 6400 in all
* Recorded onto Marantz PMD 560 Solid-state Recorder at CD samping rate, using a EV RE50 mic in a sound-dampened room.

Talkers

* 20 young adult native speakers of American English from the upper Midwest, recruited from Indiana University population

Measurements

* Selected to cover a range of acoustic attributes identified in previous works. Three sets of measures.

Measured from hand-parsed signals using Matlab:
CONS DUR = duration of consonant epoch: offset occurrence of voiced high-frequency spectra to onset of voiced high-frequency spectra
VOW DUR = duration of vowel epoch of co-syllabic /a/

Matlab scripts analyzing acoustics in consonant epoch
SPEC MEAN = mean spectral frequency of signal
SPEC VAR = standard deviation of spectral frequencies
MAX N POW = maximum noise power during cons. epoch for signal > 500 Hz
MAX VOI POW = maximum signal power for < 500 Hz
AMP DIF = difference in peak noise amplitude from average amplitude, a measure of temporal concentration of noise

Praat estimates for signal within co-syllablic /a/
F2 = Second formant centered at first voicing cycle adjacent to the consonant
DELTA-F2 = Difference in F2 between first cycle and vowel mid-point
F3 = Third formant centered at first voicing cycle adjacent to

Analysis

General Technique

* Fit models via maximum likelihood estimation according to assumptions of 3 category representation models
* Build these 3 models with 2 different assumptions concerning the generality of category internal variance.
* Model fitting (partially) with lme4 package in *R* (Bates, 2005; http://www.R-project.org).
* Use Baysian Information Criterion to compare goodness of fit of the models.

Best Estimates of Categories

SEGMENTAL model
* Each consonant indicated separately in model.
* 15 parameters - one for each consonant (-1 for intercept).

FEATURAL model
* Consonant indicated as combination of binary features
* 6 parameters - [coronal] [anterior] [sonorant] [continuant] [strident] & [voice]

ORDINAL model
* Consonant indicated by ordinal place on phonetic chart
* 3 parameters - [place], [manner], and [voice]
* [place: labial < alveolar < palato-alveolar < velar]
* [manner: nasal < stop < affricate < fricative]

Category - Internal Variance

SHARED COVARIANCE
* Model includes a single covariance matrix for the combination of each of the 11 acoustic measures.
* The combination of cells introduces 55 additional fitted parameters

SEPARATE COVARIANCE
* Model makes no assumptions about acoustic covariance across consonant categories
* Model includes a covariance matrix for all 16 consonants
* Each matrix has 55 parameters, one for each pairwise combination of 11 acoustic measures
* This introduces 917 additional fitted parameters

Evaluation

Model fits to the data employ Baysian Information Criterion (BIC). BIC balances
1) goodness of fit & 2) the complexity of the model.

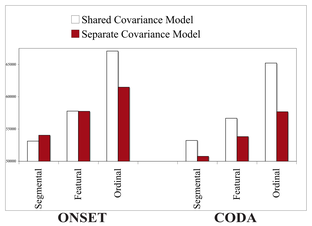
$$BIC = -2*\ln(L) + \ln(N)*k$$

(L = likelihood) (N = # of datapoints, & k = # of parameters)

Low values of BIC indicate a better model.

Results

BIC Values for Six Models



Best Estimates of Categories

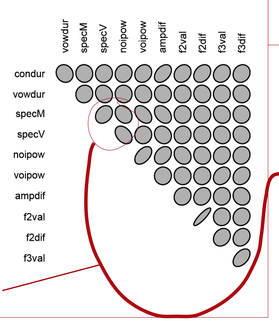
In general, BIC measures indicate better fits (lower values) for more specific models (SEGMENTAL & FEATURAL) than more general models (ORDINAL), even with added cost of additional parameters

Category - Internal Variance

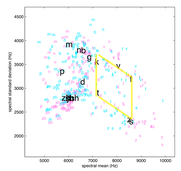
In general, BIC measures indicate better fits with separate covariance matrices, especially for codas, despite the enormous number of attendant parameters.

Why?

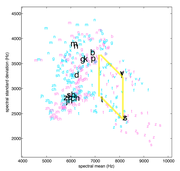
To the right are ellipse plots of every combination of measures for /b/ in onset position.
Note that the measures are generally uncorrelated.
Most ellipses are largely circular.
Hence the measures do a decent job at sampling different acoustic attributes.
To explore covariation in different segments, we focus on one corner of the overall matrix.



Onsets



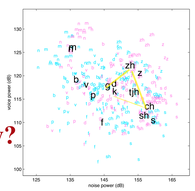
Codas

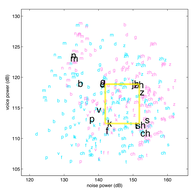


Noise spectra measures and the situation of individual subject and group means for each consonant. Some 'square' place by manner (coronal X cont) structure can be seen (in yellow).

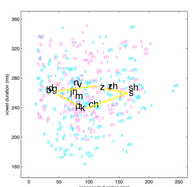
However, /p/ is misplaced in onsets, and /d/ is misplaced in codas.
Note also, ordinal place structure doesn't project from labials through coronals to dorsal

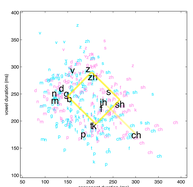
Why?



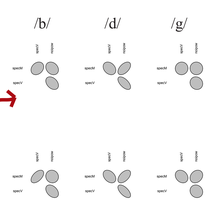


Power measures. Nice voicing X manner square in coda segments, but collapses in onsets.
Nasals & stops to the left, fricatives and affricates to the right. But not consistently.





Duration plots. Clear voicing X manner structure in coda, but affricates are wrong. Structure collapses in onsets, and nasals are misplaced.



Examples of variability across the covariance matrices

To the left are partial covariance matrices including noise power, spectral mean and standard deviation for the stops only.
Velars exhibit little covariation in the three measures, as indicated by the roughly circular plots.
Coronals exhibit negative covariance between spectral mean and variance, and positive covariance between spectral mean and noise power.
Labials, by contrast, exhibit positive covariance between spectral mean and variance, and negative covariance between spectral mean and noise power.

References

Bates, D. (2005). Fitting linear mixed models in R. *R News* 5: 27 - 30.
R Development Core Team (2008). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0. www.R-project.org.
Silbert, N.H. (under review,a). Syllable structure and integration of voicing and manner of articulation information in labial consonant identification. *JASA*.
Silbert, N.H. (under review,b). Syllable structure and integration of voicing and place of articulation information in stop consonant identification. *JASA*.
Silbert, N.H. & K. J. de Jong (2007). Laryngeal feature structure in 1st and 2nd language speech perception. In J. Trouvain and W.J. Barry (eds.), *ICPhS XVI* (pp. 1901-1905.
Silbert, N.H., & K. J. de Jong (to appear). English obstruent perception: The content of the p-map and its relationship to natural class structure. *IULC Working Papers in Linguistics*, 11.

Acknowledgements

Work supported by the NSF (grant #BCS-04406540), the NIH (R01-MH0577-17-07-A1).