LSA 511: Computational Models of Sound Change

James Kirby & Morgan Sonderegger

2 Jul 2013

Today

- Last week: what might cause merger/neutralization between perceptually adjacent categories?
- Local, similarity-based positive feedback: All else being equal, ambiguous percepts more likely to be mapped to a more active category than a less active category, even without lenition
- Feedback is local: no explicit forces keeping categories apart
- Why might we want such a thing?

Some reasons

- Contrast preservation/maintenance
- Anti-homophony effects
- Categories seem to be ‘optimally’ dispersed

Anti-homophony effects

- Sound change is more likely to be inhibited when it would result in the neutralization of a lexically or morphologically informative contrast (Martinet 1952, Blevins 1998, Kingston 2007, Blevins & Wedel 2009, Silverman 2010...)

Anti-homophony effects: Estonian

- Loss of final –n in Estonian, inhibited in NEst dialects in just those cases when it would have led to homophony between verbal inflections
- In SEst dialects, sound change took place across the board, presumably not inhibited in this same context b/c retention of –ʔ meant that the verbal forms could still be distinguished:

Northern Estonian	*Southern Estonian*	*Proto-Balto-Finnic*	
kannan	kanna	*kanna-n	'I carry'
kanna	kannaʔ	*kanna-ʔ	'Carry!'

Table: Estonian verb forms after loss of ʔ and *n*. After Campbell (1998).

Anti-homophony: Banoni (Lincoln 1976, Blevins & Wedel 2009)

Proto-Oceanic	Banoni	gloss
*pupu	vuu ∼ vu	'fish trap'
*paqoRun	voom ∼ vom	'new'
*poñu	vom	'turtle'

bare N	1sg poss	gloss
tama	tamaa	'father/my father'
punu	punuu	'hair/my hair'
kasi	kasii	'brother/my brother'

'Variant trading' (Wedel 2004, Blevins & Wedel 2009)

- Exemplar model w/ fixed-length lists, Gaussian noise
- Likelihood that target is labeled category c proportional to distance to category mean
- Variant trading: outputs are always stored back as their intended category (English 'cook' vs. 'cooked')
- `http://dingo.sbs.arizona.edu/~wedel/simulations/1DVariantTrading.html`

'Variant trading': Questions

- Does the lack of activation weighting matter?
- How about the fixed length lists?
- What is the necessary relation of word to subword representations here?
- Does the P model display this kind of behaviour? Should it?

‘Variant trading’: Assessment

- Suggests a slightly more complex knowledge state...
- ... as well as assumptions about communication
- Learning strategy? Qualitative predictions?

Dispersion

- An observation: phoneme inventories seem to like to be dispersed evenly throughout the perceptual space.

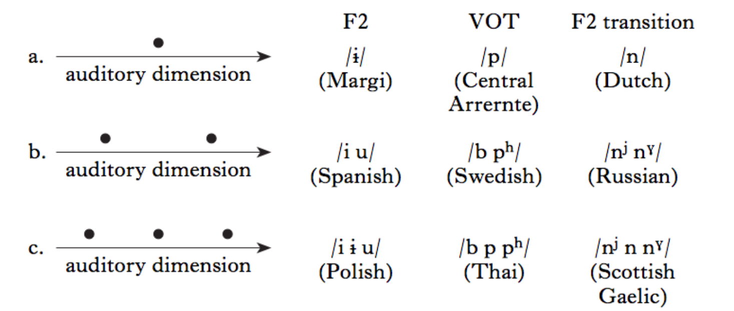


Figure 1
Observed auditory dispersion.

from Boersma & Hamann (2008).

Typology of dispersion

- Preference for center (if 1 category, then centered)
- Excluded center (if 2 categories, then not centered)
- Equal distances (categories prefer to be perceptually distinct)
- Larger inventory → larger space
- Fewer categories → more variation
- Chain shifts

Explaining dispersion

- Ohala (1981), Blevins (2004), etc.: sound change is caused by reanalysis of imperfectly transmitted (perceived) sounds
- ‘Sound change through misperception ... can only hope to account for neutralization, not dispersion or enhancement’ (Flemming 2005:173)
- Vanilla OT markedness & faithfulness constraints can’t account for ‘excluded center’ effects (*[ɨ] >> *[u] > *[i])

Goal-oriented: Dispersion contraints

	F2 6	5	4	3	2	1	F1
	i	i̠	y	ɨ	ɯ	u	1
		ɪ	ʏ			ʊ	2
		e̝	ø̝		ɤ̝	o̝	3
		e	ø	ə	ɤ	o	4 F1
			ɛ	ɜ	ʌ	ɔ	5
			æ	ɐ		ɑ	6
				a	a̠		7

		MINDIST = F1:2	MINDIST = F1:3	MAXIMIZE CONTRASTS	MINDIST = F1:4	MINDIST = F1:5
a.	i-a			✓✓!		
b.	☞ i-e-a			✓✓✓	**	**
c.	i-e̝-ɛ-a		*!**	✓✓✓✓	***	*****

From Flemming, 2001

Is there another way?

- Along with Padgett (2001), Sanders (2003), these works evaluate not input forms, but entire inventories or languages
- But: what if dispersion constraints, while expressing surface-true observations about sound systems, are epiphenomenal of an underlyingly non-goal-oriented mechanism (Padgett 2003)?

Boersma & Hamann 2008

- Want to explain auditory dispersion 'without resorting to exemplar theory' (218)
- Instead, use Boermsa's bidirectional model (with some notion of articulatory ease)
- 'Observationally optimising but underlyingly non-telelogical'

Framework: Bidirectional phonetics

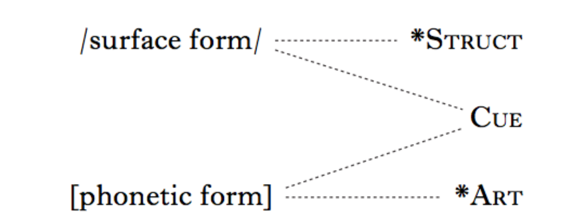


Figure 3
Grammar model for phonetics.

- Same grammar is used in both production and comprehension
- Why model phonetics with constraints?

Framework: Bidirectional phonetics

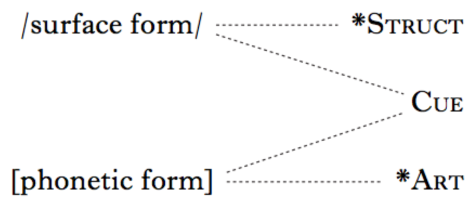


Figure 3
Grammar model for phonetics.

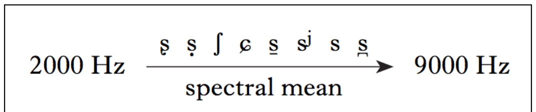
- ‘The output of the perception process tends to be restricted by the same structural constraints that have been proposed for phonological production’ (227)

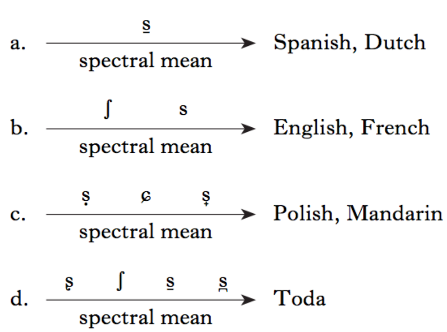
Types of constraints

- STRUCT constraints (e.g. ONSET): evaluate phonological form only
- CUE constraints (e.g. *[long vowel duration] /obs, −voice/): evaluate relation between phonetic and phonological form
- ART constraints (e.g. *31 Erb): evaluate phonetic form only
- Here, mainly concerned with the CUE and ART constraints

Case studies: sibilant dispersion

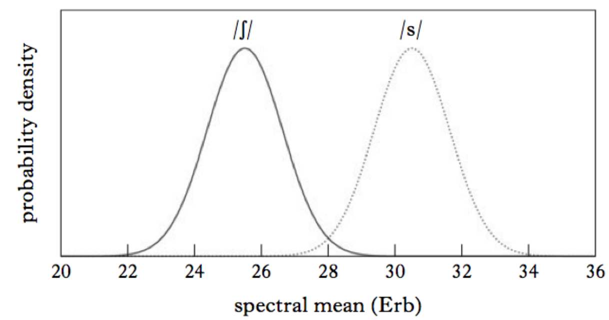
Primary acoustic continuum: spectral center of gravity or spectral mean





Learning a perceptual grammar

- Given a fixed distribution of tokens along an acoustic dimension...



Learning a perceptual grammar

- ...learn the optimal ranking of cue constraints for that distribution:

(2) *A perception tableau for classifying tokens with a spectral mean in English*

[26·6 Erb]	*[26·5]/s/	*[26·6]/s/	*[26·7]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/
a. /s/		*!				
☞ b. /ʃ/					*	

Lexicon-driven perceptual learning

- Learner optimally re-ranks cue constraints consistent w/input, via the GRADUAL LEARNING ALGORITHM (Boersma & Hayes 2001)
- All constraints favouring the correct category are moved up, and those favouring the incorrect category are moved down:

(3) *A learner's perception tableau with reranking of cue constraints*

[26·6 Erb]	*[26·5]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/	*[26·6]/s/	*[26·7]/s/
☞ a. /s/					← *	
✓ b. /ʃ/			*! →			

Lexicon-driven perceptual learning

- Learners come to display maximum-likelihood behaviour with respect to unambiguous inputs...(what is this?)
- ...and probability match with respect to ambiguous inputs (i.e. those that fall in the overlap between the two distributions)

(3) *A learner's perception tableau with reranking of cue constraints*

[26·6 Erb]	*[26·5]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/	*[26·6]/s/	*[26·7]/s/
☞ a. /s/					← *	
✓ b. /ʃ/			*! →			

Lexicon-driven perceptual learning

- These simulations crucially assume random prelexical perception but perfect lexical storage
- Learners know how many sibilants the language has, but not where their distributions lie within the frequency continuum.

(3) *A learner's perception tableau with reranking of cue constraints*

[26·6 Erb]	*[26·5]/s/	*[26·7]/ʃ/	*[26·6]/ʃ/	*[26·5]/ʃ/	*[26·6]/s/	*[26·7]/s/
☞ a. /s/					← *	
✓ b. /ʃ/			*! →			

Demo: English perception

- Start with 322 equally ranked cue constraints
- 1m data points drawn randomly from each distribution
- Rerank constraints if perceived category $\neq$ lexical category

Demo: English perception

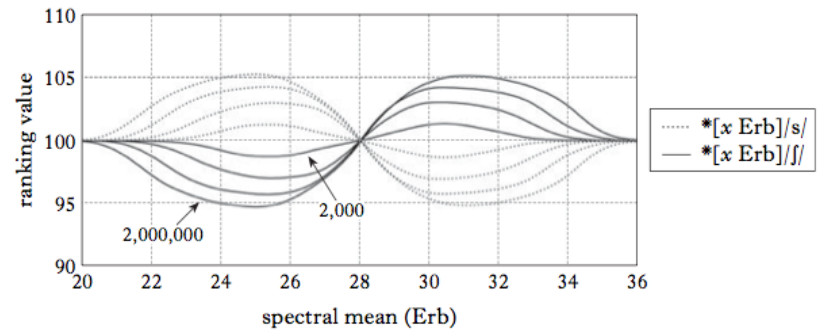


Figure 7

A virtual learner acquiring the perception of the two sibilant categories of English: perception grammars after 2,000, 20,000, 200,000 and 2,000,000 input tokens.

Demo: English perception

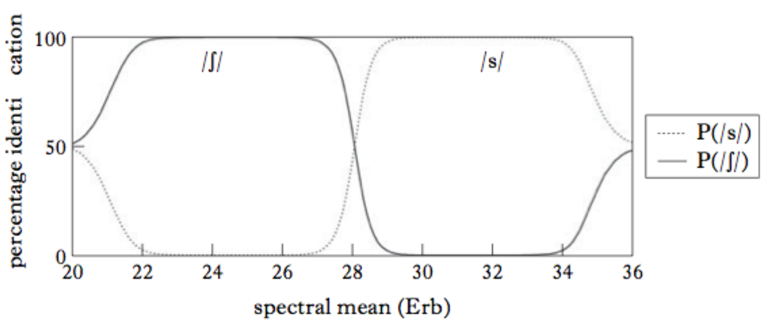


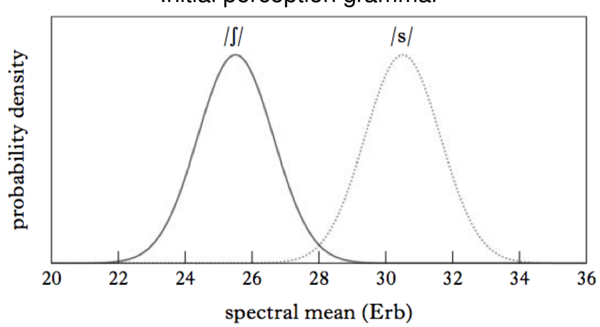
Figure 8

Identification curves for the virtual listener of Fig. 7 (final state).

A side-effect: prototypes

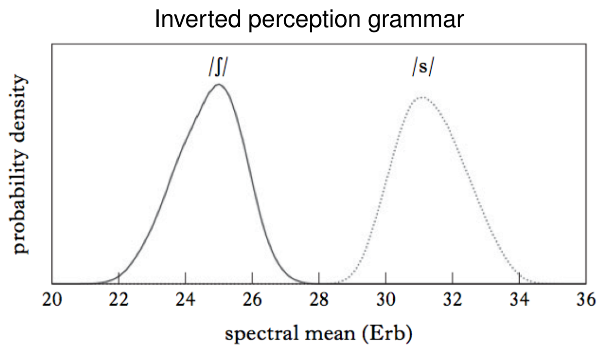
- When just cue constraints are involved, listener-turned-talker tends to prefer to produce tokens at the periphery:

Initial perception grammar



A side-effect: prototypes

- When just cue constraints are involved, listener-turned-talker tends to prefer to produce tokens at the periphery:



A side-effect: prototypes

- This is because production is based on sampling from the learned perception grammar (with evaluation noise) rather than from the distribution of actually learned examples

(4) *A production tableau with cue constraints only*

/s/	*30·6 /s/	*30·7 /s/	*30·8 /s/	*31·5 /s/	*30·9 /s/	*31·4 /s/	*31·3 /s/	*31·0 /s/	*31·2 /s/	*31·1 /s/
a. [30·6 Erb]	*!									
b. [30·7 Erb]		*!								
c. [30·8 Erb]			*!							
d. [30·9 Erb]					*!					
e. [31·0 Erb]								*!		
☞ f. [31·1 Erb]										*
g. [31·2 Erb]									*!	
h. [31·3 Erb]							*!			
i. [31·4 Erb]						*!				
j. [31·5 Erb]				*!						

*Art(iculatory) constraints

- By hypothesis, production involves more than cue constraints.

(5) *A production tableau with cue constraints and articulatory constraints*

/s/	*31·2	*31·1	*31·0	*30·9	*30·6 /s/	*30·8	*30·7 /s/	*30·7	*30·8 /s/	*30·6	*30·9 /s/	*31·0 /s/	*31·2 /s/	*31·1 /s/
a. [30·6 Erb]					*!					*				
☞ b. [30·7 Erb]							*	*						
c. [30·8 Erb]						*!			*					
d. [30·9 Erb]				*!							*			
e. [31·0 Erb]			*!									*		
f. [31·1 Erb]		*!												*
g. [31·2 Erb]	*!												*	

- B&H assume a relationship between auditory peripherality and articulatory effort: the more peripheral, the harder to produce.

*Art(iculatory) constraints

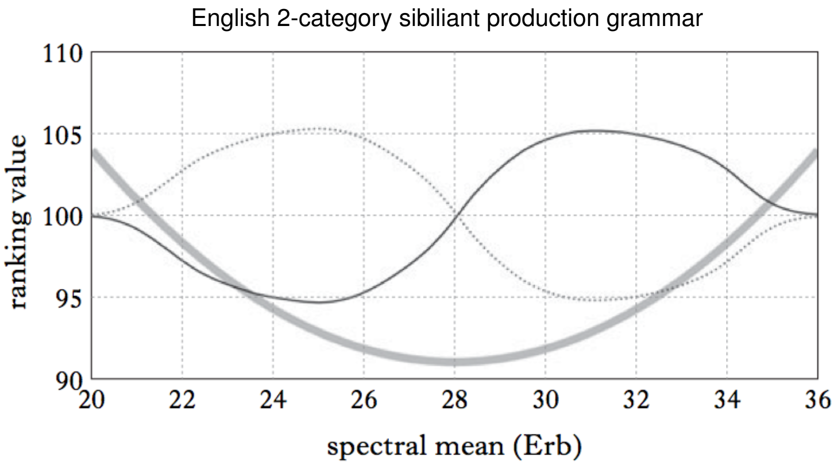
- Modeled by (universally ranked) Art(iculatory) constraints:

(5) *A production tableau with cue constraints and articulatory constraints*

/s/	*31·2	*31·1	*31·0	*30·9	*30·6 /s/	*30·8	*30·7 /s/	*30·7	*30·8 /s/	*30·6	*30·9 /s/	*31·0 /s/	*31·2 /s/	*31·1 /s/
a. [30·6 Erb]					*!					*				
☞ b. [30·7 Erb]							*	*						
c. [30·8 Erb]						*!			*					
d. [30·9 Erb]				*!							*			
e. [31·0 Erb]			*!									*		
f. [31·1 Erb]		*!												*
g. [31·2 Erb]	*!												*	

- Notice how the Cue constraints can be re-ranked among the Art constraints.

*ART(iculatory) constraints



Simulating sound change: English

- Parameters:
  - transmission noise, evaluation noise, number of examples, decision strategy
- Stable English
- Exaggerated English
- Skewed English

Simulating sound change: Polish

- Old Polish: asymmetric sibilant inventory /ʃ sʲ s/ (Carlton 1991)
- Modern Polish: basically symmetric sibilant inventory /ʂ ɕ s̪/
- Around the 13th century, [sʲ] > [ɕ]...
- ...300 years later, [ʃ] > [ʂ]
- Polish (decision strategies, transmission noise, 10 examples...)

Discussion

- 'Doomed to success'?
- Where do cue constraints come from?
- Does the B&H model make different predictions from exemplar-theoretic models?
- Population dynamics?
- What about merger? How could this be implemented?

Standard Eastern Norwegian

- Starting from CScand /s̠ ʃ s̢/, predicts [s̟ ɕ ṣ] instead of [s̟ ç ṣ]
- Fails to predict attested mergers of /s̢/, /ʃ/ > [ṣ] and /s̢/, /ç/ > [ç]

From Williams (2012).

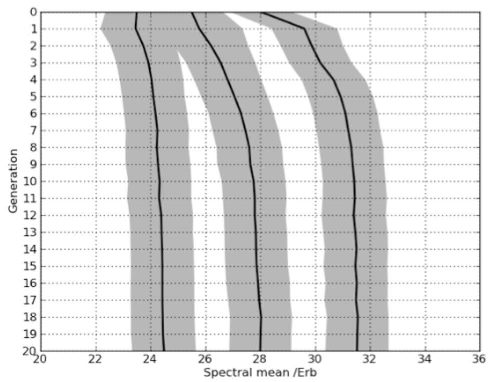


Figure 3
The evolution of the three initial standard eastern Norwegian sibilants over 20 generations.

(B&H's) summary

- Optimal dispersion effects can emerge 'innocently'
- Bidirectionality predicts listener-oriented effects
- Suggests a larger role for general constraint-based theories of language processing

Our summary

1. Knowledge state of individuals
2. Network/social structure
3. Assumptions about communication
4. Learning algorithm

What have we learned?

- (Non-)existence proof(s)?
- Explicitness?
- Counterintuitive results?
- Qualitative predictions?
- Baseline?