Phonetic and phonological factors in coronal-to-dorsal perceptual assimilation

Eleanor Chodroff and Colin Wilson

Johns Hopkins University

Laboratory Phonology 2014 | Tokyo, Japan

Perceptual Assimilation

Listeners often identify non-native sounds and sequences as instances of native structures / fail to discriminate foreign and native structures

Norwegian [y] → English [i] at a rate of .90+

French [ebdo] → Japanese [ebɯdo] at a rate of .60+

Two factors are known to influence patterns of perceptual assimilation

- Acoustic-phonetic (auditory) similarity
- Phonological constraints and processes

What are the relative contributions of acoustic similarity and phonology in accounting for detailed patterns of assimilation?

Coronal-to-Dorsal Perceptual Assimilation

French and American English listeners often misperceive Modern Hebrew coronal-lateral clusters as beginning with dorsal stops

	Fr ident*	AE ident*
MH *tl* → *kl*	.81	.86
MH *dl* → *gl*	.29	.39

*Hallé & Best, 2007

- Other perceptual repairs (e.g., epenthesis, coronal-to-labial) found rarely
- Asymmetry between *tl* and *dl* puzzling on typological grounds
- Acoustic-phonetic account not strongly supported by Hallé et al. analysis

Outline

1 Experiment 1a: Laboratory Perception – MH Speaker 1

2 Experiment 1b: MTurk Perception – MH Speaker 1

3 Experiment 2: MTurk Perception – Additional 3 MH Speakers

4 Modeling the perceptual findings
   i. English productions and acoustic analysis
   ii. Phonetic likelihood model
   iii. Bayesian model with phonetic likelihood & phonotactic prior

Procedure

Procedure adapted from studies by Hallé et al.

Stimuli:

- Female native MH talker recorded stimuli in frame context from prompts presented in Hebrew orthography

  $t\ d\ k\ g \times ʁ\ l \times i\ e\ a\ o\ u \times 4$

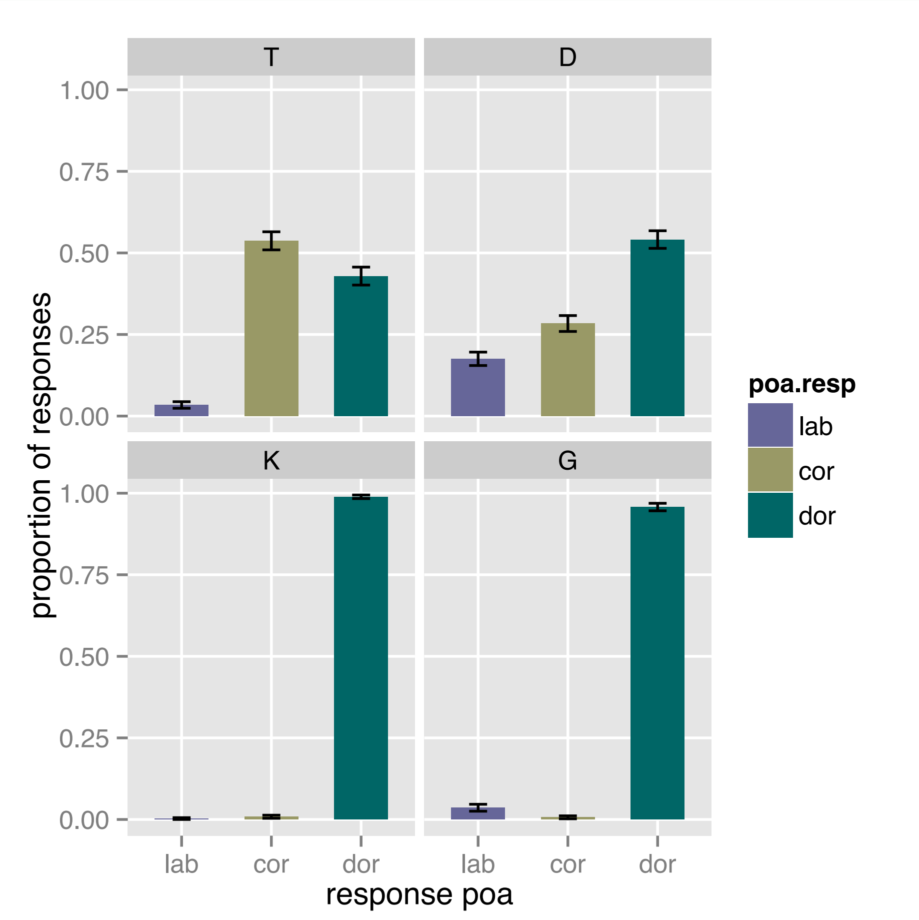
- 8 items removed due to poor recording or unclear production

Task:

- 18 AE listeners in sound-attenuated booth heard each stimulus twice consecutively, with item order randomized across participants, and identified the initial consonant as P T K B D G
- Subsequent to identification each item was presented again for goodness rating, but rating results not reported here

Results

pre-l response pattern



pre-l accuracy: 69.1%
pre-ʁ accuracy: 98.1%

Logistic mixed-effects analysis of **place perception accuracy**

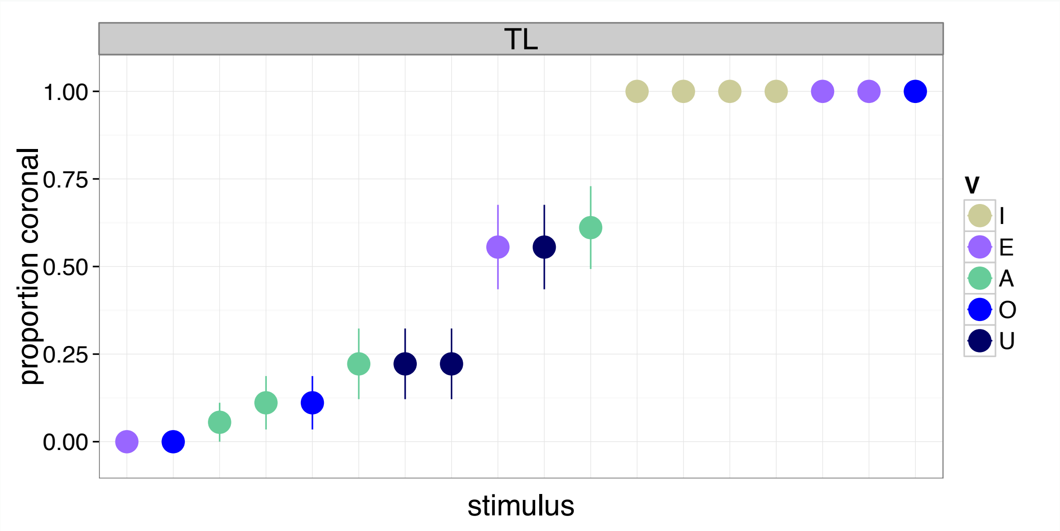
	β estimate	*p*-value
(intercept)	4.85	<0.001
poa	-1.86	<0.001
voice	0.91	<0.01
C2	-1.87	<0.001
poa:voice	0.01	0.96
poa:C2	**-1.72**	**<0.001**
voice:C2	0.16	0.56
poa:voice:C2	0.10	0.68

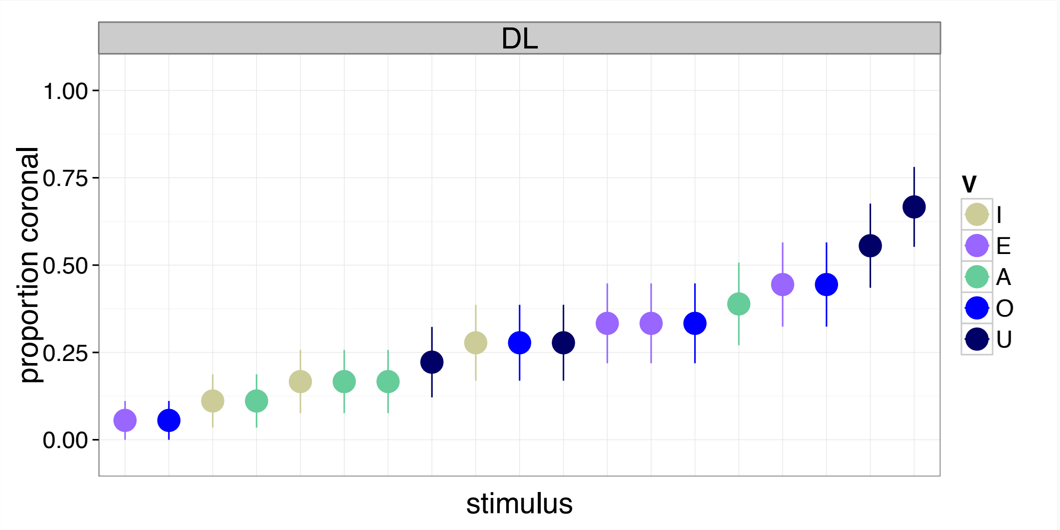
poa (cor 1 vs dor -1), voice (vcl 1 vs vcd -2), C2 (lateral 1 vs rhotic -1)
*analyzed with random intercepts for participant and item

- less accurate with coronals
- more accurate with voiceless stops
- less accurate with the coronal-lateral cluster

Stimulus-specific pattern

Experiment 1a: Lab Perception

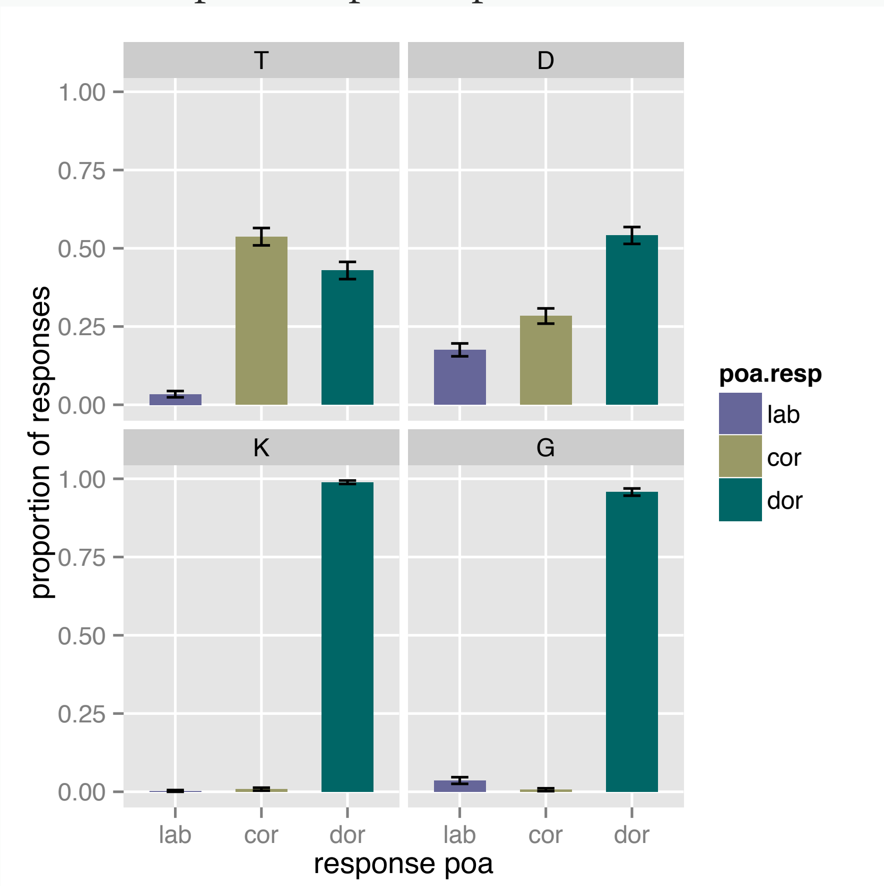




Outline

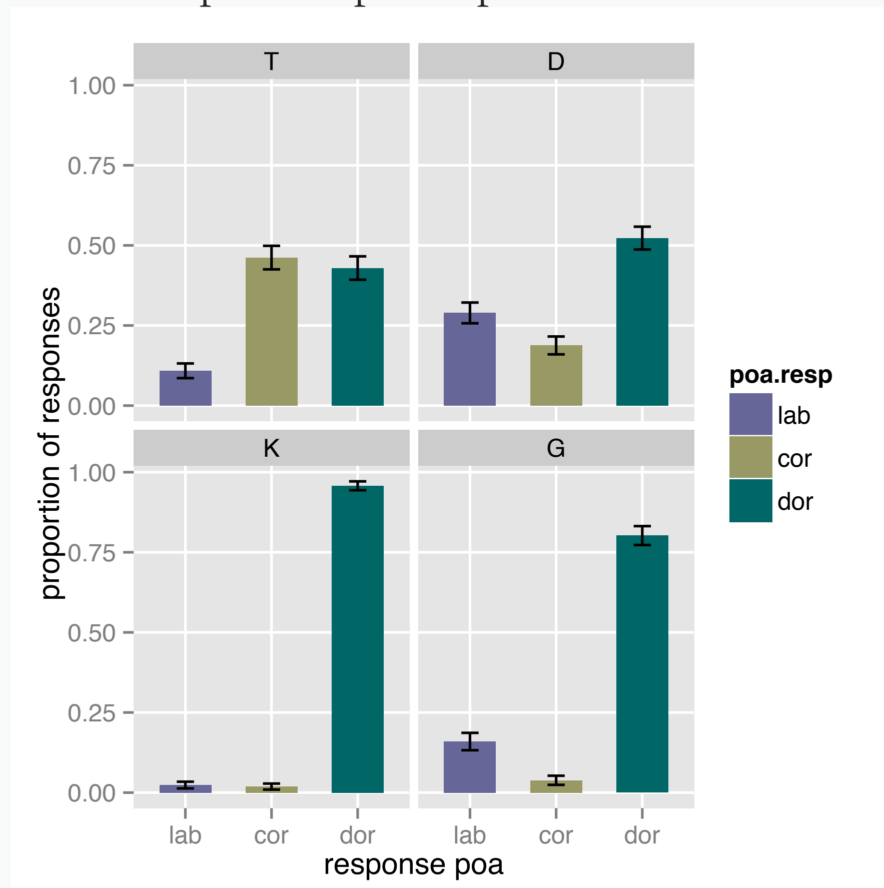
MTurk Replication

F1 Laboratory
pre-l response pattern



pre-l accuracy: 69.1%
pre-ʁ accuracy: 98.1%

F1 MTurk
pre-l response pattern



pre-l accuracy: 60.8%
pre-ʁ accuracy: 90.7%

MTurk Replication

Strong correlation between stimulus-specific coronal response rates in lab and MTurk experiments:

- all stimuli: $r = 0.96$
- tl, dl stimuli: $r = 0.89$

Same pattern of significance as in the laboratory experiment

Logistic mixed-effects analysis of **place perception accuracy**

	β estimate	p-value
(intercept)	3.07	<0.001
poa	-1.87	<0.001
voice	1.01	<0.001
C2	-1.74	<0.001
poa:voice	-0.38	0.06
poa:C2	**-0.67**	**<0.001**
voice:C2	0.26	0.18
poa:voice:C2	0.03	0.87

poa (cor 1 vs dor -1), voice (vcl 1 vs vcd -2), C2 (lateral 1 vs rhotic -1)

*analyzed with random intercepts for participant and item

Outline

1 Experiment 1a: Laboratory Perception – MH Speaker 1

2 Experiment 1b: MTurk Perception – MH Speaker 1

3 Experiment 2: MTurk Perception – Additional 3 MH Speakers

4 Modeling the perceptual findings

   i. English productions and acoustic analysis
   ii. Phonetic likelihood model
   iii. Bayesian model with phonetic likelihood & phonotactic prior

Additional Speakers

Stimuli:

- One additional female and two male native MH talkers recorded stimuli in frame context from prompts presented in Hebrew orthography

  t d k g × *ʁ l* × *i e a o u* × 4-5

- 4 recordings per type

Task:

For each speaker:

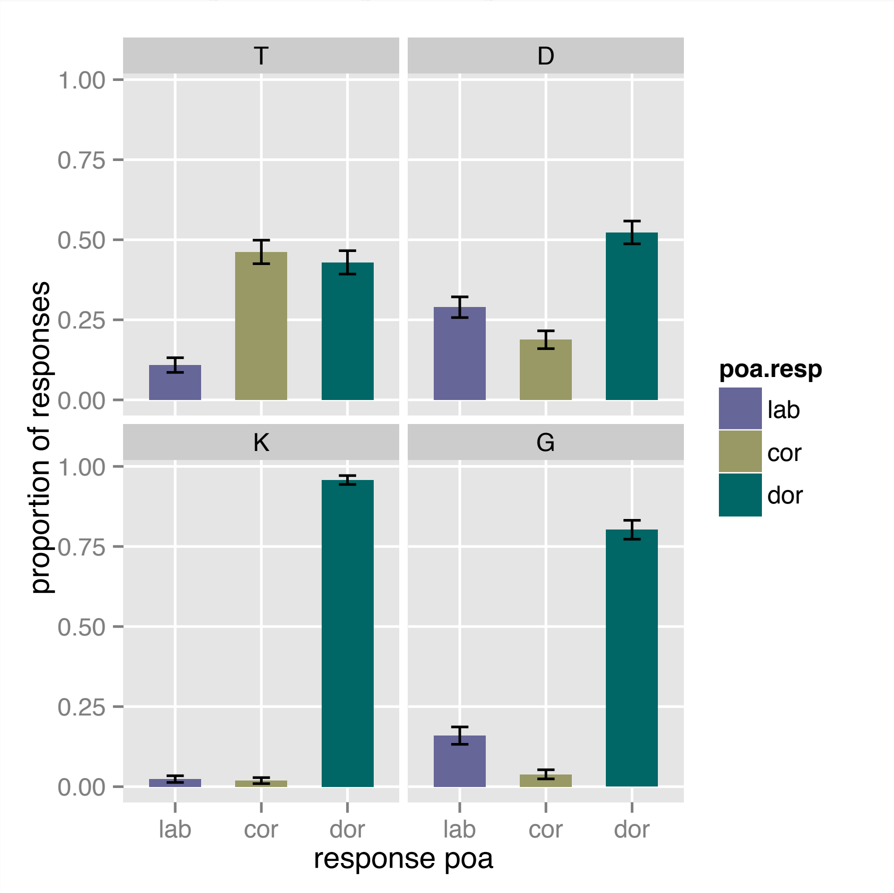
- 20 AE listeners heard each stimulus twice consecutively, with item order randomized across participants, and identified the initial consonant as P T K B D G

Talker Differences

Experiment 2: MTurk – Additional Speakers

F1 MTurk

pre-l response pattern

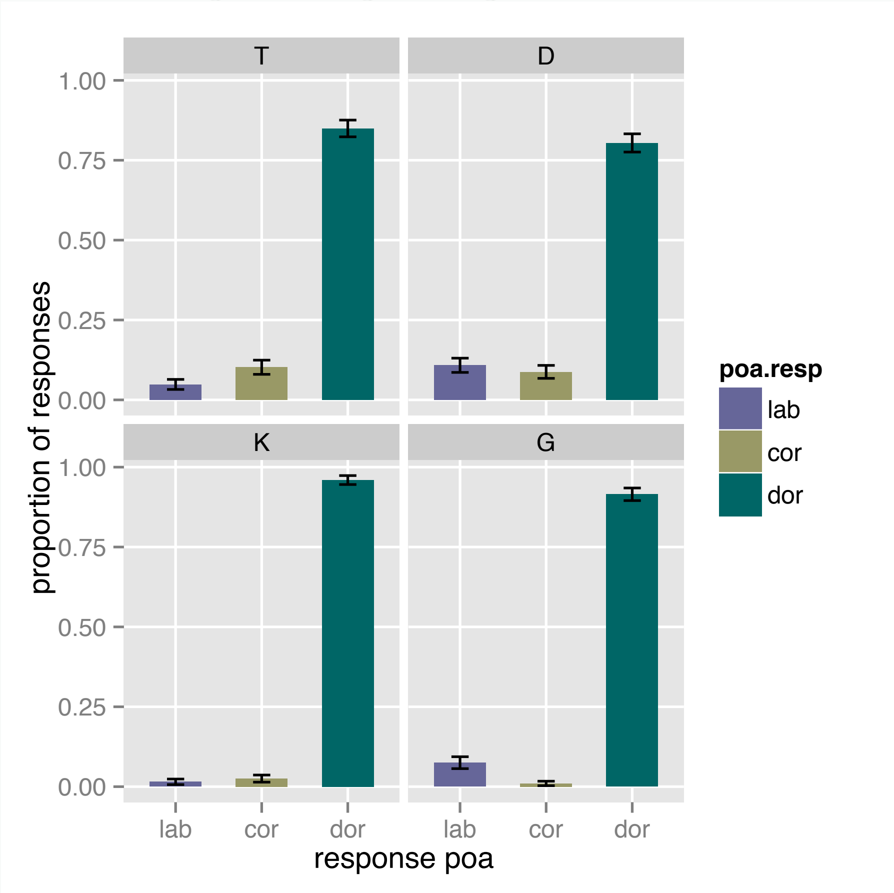


pre-l accuracy: 60.8%

pre-ʁ accuracy: 90.7%

M2 MTurk

pre-l response pattern

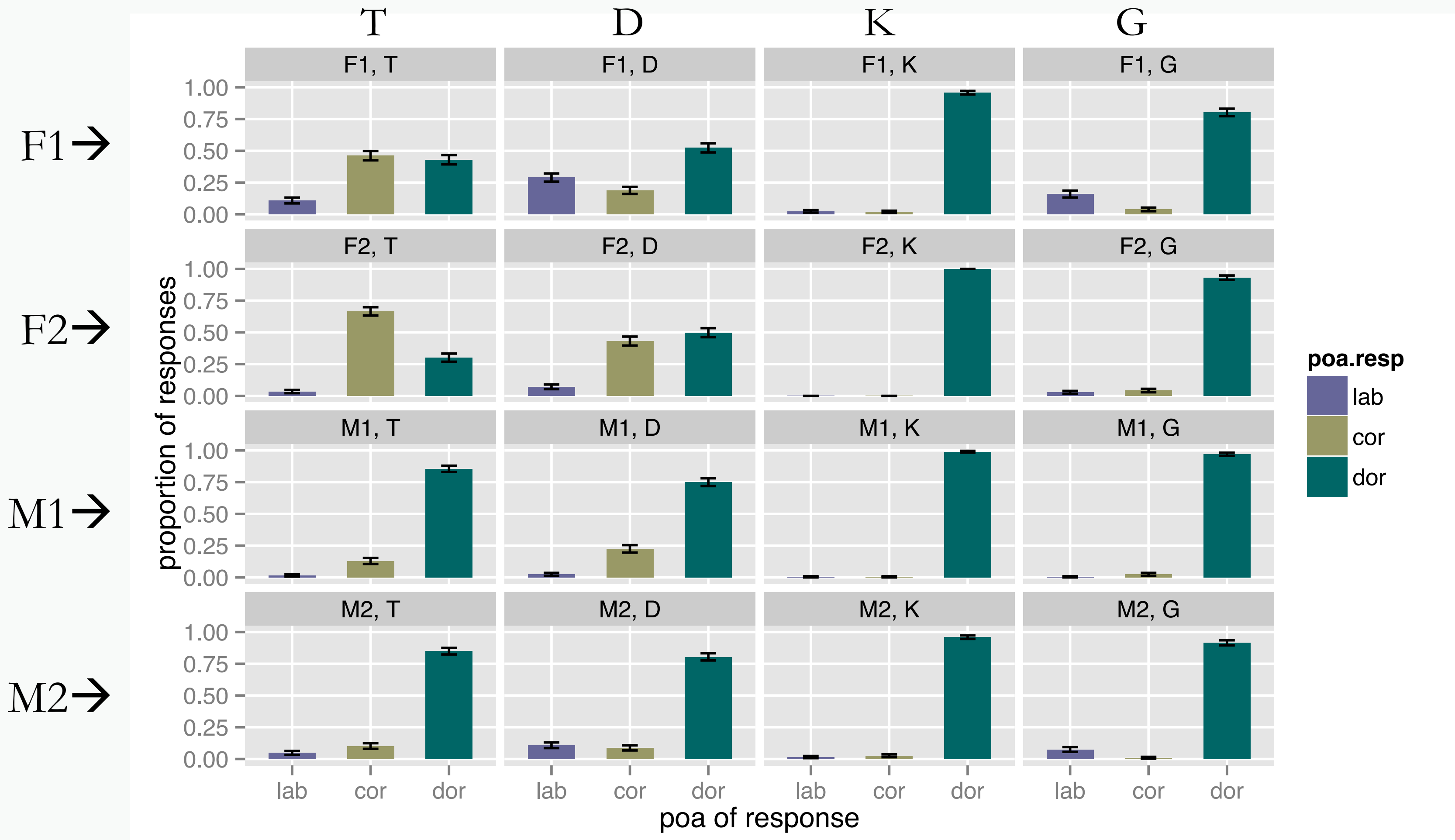


pre-l accuracy: 52.6%

pre-ʁ accuracy: 91.1%

Results

pre-l response pattern



pre-l accuracy range: 52.6% (M2) – 76.2% (F2)

pre-ʁ accuracy range: 90.7% (F1) – 98.2% (M1)

Results

Logistic mixed-effects analysis of
place perception accuracy

	β estimate	*p*-value
(intercept)	2.48	<0.001
poa	-1.52	<0.001
voice	0.75	<0.001
C2	-1.43	<0.001
talkerF2	2.35	0.80
talkerM1	1.15	<0.01
talkerM2	-0.54	0.15
poa:voice	-0.28	<0.05
poa:C2	**-0.52**	**<0.001**
voice:talkerM1	-0.53	<0.05
C2:talkerM1	-0.77	<0.05
poa:C2:talkerF2	-1.59	0.86
poa:C2:talkerM1	**-1.35**	**<0.001**
poa:C2:talkerM2	**-1.05**	**<0.001**

poa (cor 1 vs dor -1), voice (vcl 1 vs vcd -2), C2 (lateral 1 vs rhotic -1), talker (F1 0 vs F2 1; F1 0 vs M1 1, F1 0 vs M2 1)

*analyzed with random intercepts for participant and item

Selected effects and interactions
↵

Includes results from MH Speaker 1 MTurk perception

- less accurate with coronals
- more accurate with voiceless stops
- less accurate with lateral liquid
- less accurate with coronal-lateral clusters
- less accurate with coronal-lateral clusters for M1 and M2

Interim Summary

Coronal-to-dorsal perceptual assimilation observed for a large set of stimuli (~700, 175 critical) from multiple talkers

cf. 24 critical stimuli from one male talker in Hallé & Best (2007)

Rate of coronal perception and voiceless-voiced asymmetry varies greatly across talkers and across stimuli within talkers

M vs. F talker difference is strong but confounded

Remaining Questions:

- Can acoustic-phonetic properties of the stimuli account for the perception results?
- Specifically, how good are the Hebrew stop consonants as examples of English stop consonants?
- What is the role of phonological bias in perceptual assimilation?

Outline

1 Experiment 1a: Laboratory Perception – MH Speaker 1

2 Experiment 1b: MTurk Perception – MH Speaker 1

3 Experiment 2: MTurk Perception – Additional 3 MH Speakers

4 Modeling the perceptual findings
   i. English productions and acoustic analysis
   ii. Phonetic likelihood model
   iii. Bayesian model with phonetic likelihood & phonotactic prior

English productions and acoustics

English corpus of CVC syllables

p b t d k g × *i ɪ e ɛ æ ʌ a ɔ o u* × *t* × 5

18 speakers (4 male)

Also recorded CLVC dorsal-initial syllables for the same speakers (not used for model training)

Resampled at 16kHz, high-pass filtered at 100Hz, pre-emphasized from 1000Hz

(Hallé & Best, 2007; Sundara, 2005)

Acoustic-Phonetic Measures:

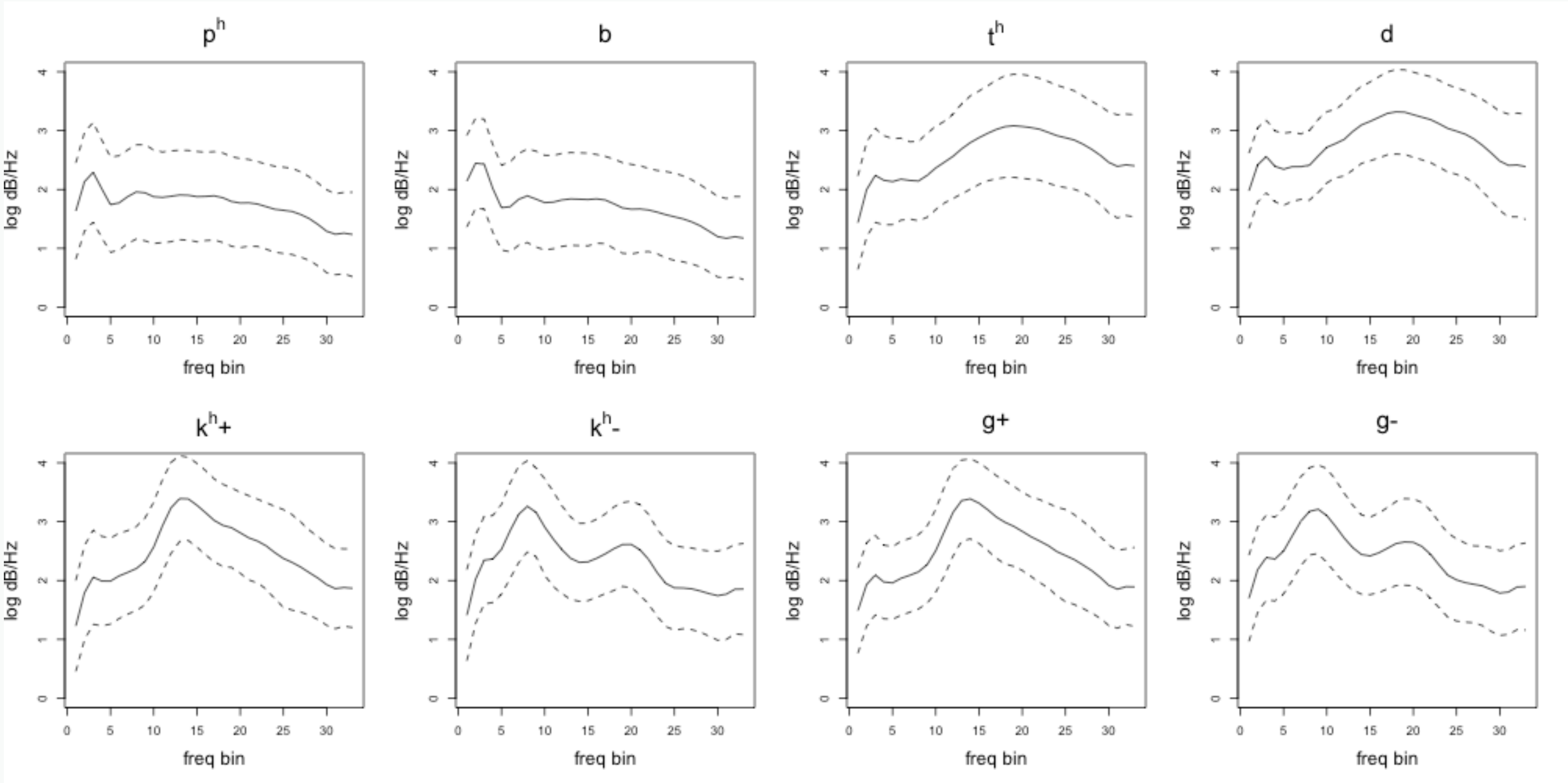
Spectral shape of the initial burst release (~ 8.5ms)

- Computed DFT for 7 consecutive 3ms Hamming windows, shifted 1ms apart, first window centered on burst release (Hanson & Stevens, 2003)
- 33-bin smoothed spectrum created by averaging power within each bin across all windows

Also measured F2 onset and trajectory of the following vowel, amplitude of the initial 10ms burst relative to following sonorant, stop burst duration — but these did not substantially improve predictions of *stop place* perception.

Phonetic likelihood model

Multidimensional Gaussian distributions fit to the smoothed spectra (and total log power) of eight English stop allophones



Maximum likelihood predictions of *stop place*: 91% correct on CVC (training data), 88% on CLVC (productions from same English speakers)

Phonetic likelihood model

Smoothed spectra (and log power) of Hebrew stimuli measured in the same way as English and stop place of each stimulus classified by max. likelihood

$$\text{predicted-place}(\text{trial}_i) = \text{PLACE}[\arg\max_x p(\text{stim}_i \mid x)]$$
$$\text{where } x \in \{ p^h, b, t^h, d, k^h+, k^h-, g+, g- \}$$

Talker	Chance	Phonetic model	
		C{L,R}V	CLV
F1 (n = 2736 | 1601)	33% –3005 | –1758	75% | 70% –1787 | –1331	69 % | 64 %
F2 (n = 1601)	33% –1758	73% –1090	66%
M1 (n = 1601)		79% –902	72%
M2 (n = 1601)		63% –1272	49%

Bayesian model

Assess the contribution of phonology (phonotactics) by combining acoustic likelihood with a perceptual prior according to Bayes' Theorem

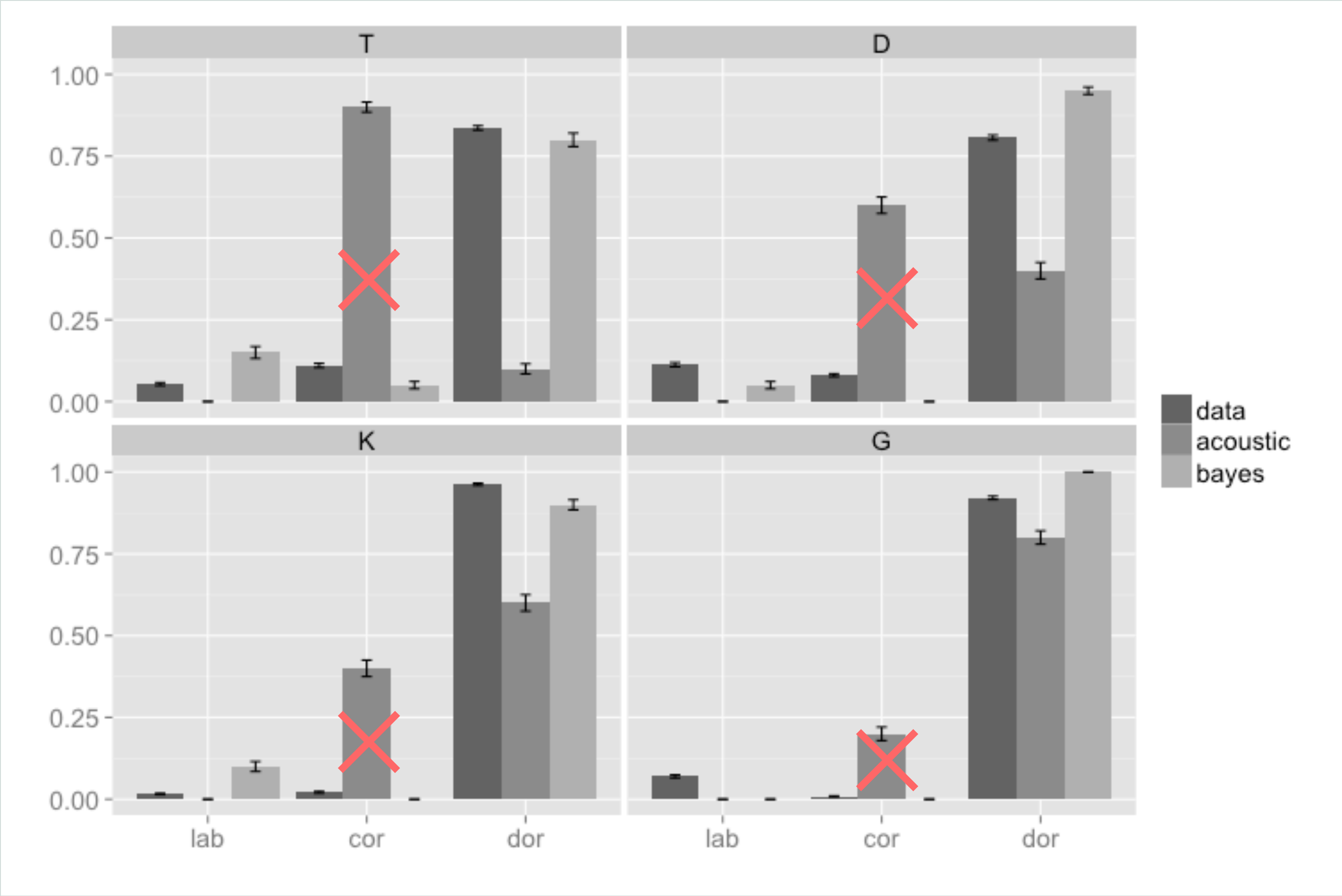
predicted-place(trial$_i$) = PLACE[arg max$_x$ p(stim$_i$ | x) · p(x | **approximant$_i$**)]
where $x \in$ { p^h, b, t^h, d, k^h+, k^h-, g+, g- }

Talker	Chance	Phonetic model		Bayesian model	
		C{L,R}V	CLV	C{L,R}V	CLV
F1 (n = 2736 | 1601)	33% –3005 | –1758	75% | 70% –1787 | –1331	69 % | 64 %	79% | 72% –1679 | –1266	77% | 69%
F2 (n = 1601)	33% –1758	73% –1090	66%	74% –1042	68%
M1 (n = 1601)		79% –902	72%	85% –738	84%
M2 (n = 1601)		63% –1272	49%	80% –1020	83%

Bayesian model

Phonotactic contribution to perception of stimuli from talker M2

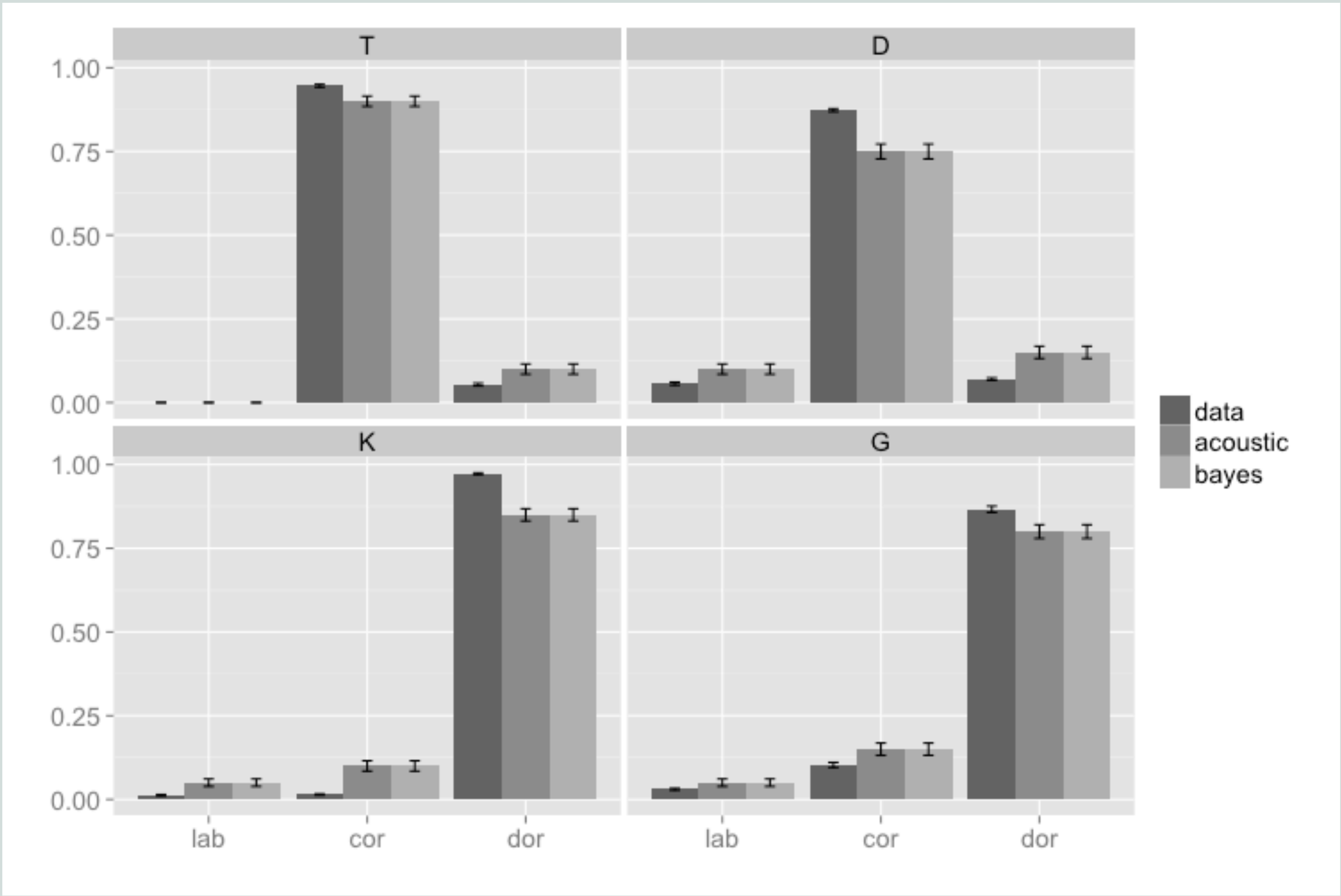
CLV stimuli



Bayesian model

Phonotactic contribution to perception of stimuli from talker M2

CRV stimuli



Summary

- Perception of the same nonnative cluster types varies across talkers (and across stimuli within talker), extending previous cross-language comparisons (Best & Hallé, 2011).

- Cross-language perception models should provide quantitative accounts of responses to individual talkers (stimuli), and more general patterns, in terms of native knowledge.

  (see also Wilson & Davidson, 2013; Wilson, Davidson, & Martin (to appear) for related developments; additionally, Strange et al., 2005; Escudero et al., 2012 for acoustic classification)

Summary

- Formally characterizing **phonetic similarity (likelihood)** w.r.t. native language is logically necessary for perception models and results in high performance
  - English phonetic model alone predicts 63% – 79% (49% – 72%) of *trial-level data* (place identifications) in the current experiments with no fit parameters
  - Phonetic likelihood has a straightforward relationship to talker / stimulus variability and provides a baseline against which more complex models can be assessed
  - Phonetic models can be extended to incorporate further cues (including dynamic transitions), multiple mixture components (sub-allophones), listener differences, …

- **Phonotactic knowledge** can be formally integrated with phonetic similarity using Bayes' Theorem, and doing so does improve measures of model fit (72% – 85%, 68% – 84%)

Acknowledgments

Modern Hebrew Speakers GE, SM, YM, and ZC

Undergrad RAs Anthony Arnette and Samhita Ilango

NYU Phonetics and Experimental Phonology Lab

NSF grant BCS-1052784 to Colin Wilson