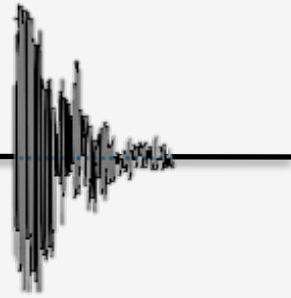
Burst Spectrum as a Cue to Stop Consonant Voicing

English Production and Perception Results

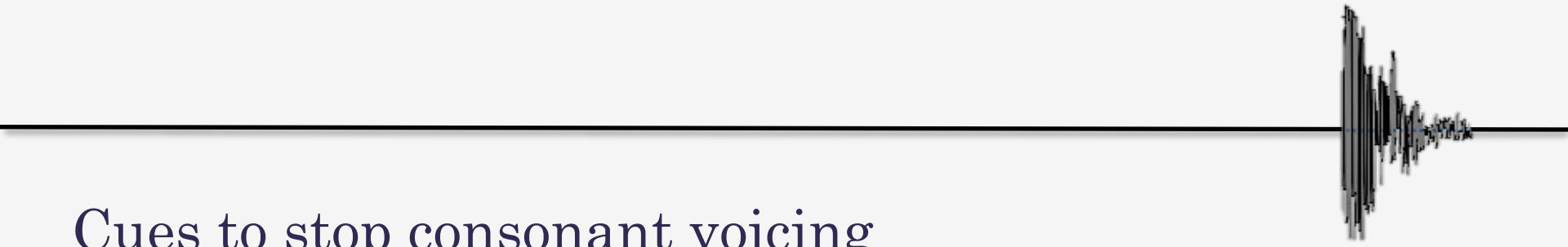
Eleanor Chodroff and Colin Wilson
Johns Hopkins University



Summerfield and Haggard (1977), Lisker (1978), Repp (1979), Lisker (1986)

voice onset time
F1 onset
F1 transition
F_0 contour
relative amplitude of aspiration
following vowel duration

spectral shape of the burst:
lower frequencies for *voiced* stops

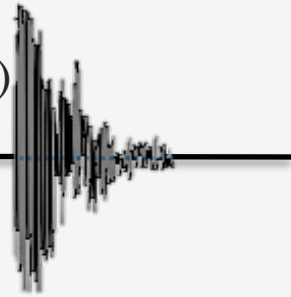


Cues to stop consonant voicing

"Since most of our lax [voiced] stops were pronounced with vocal-cord vibration, their spectra contained a strong low-frequency component…

The lax stops also show a significant drop in level in the high frequencies. This high-frequency loss is a consequence of the lower pressure associated with the production of lax stops and is therefore a crucial cue for this class of stops."

Halle, Hughes, and Radley (1957)



Background: Production

labials		
/p/	/b/	Δ
1910	1163	747❖
		Hz

coronals		
/t/	/d/	Δ
3600	3300	300+
5649	5225	424❖
4900	4400	500◆
		Hz

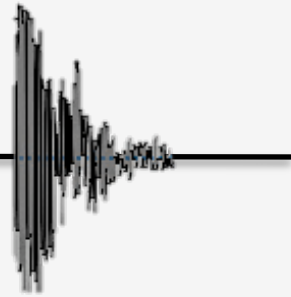
dorsals		
/k/	/g/	Δ
1940	1910	30+
2261	2268	-7❖
		Hz

+ = Zue (1976) using peak frequency

❖ = Parikh and Loizou (2005) using peak frequency

◆ = Sundara (2005) using mean frequency (CoG)

see also Van Alphen and Smits (2004), Vicenik (2010), Kirkham (2011)



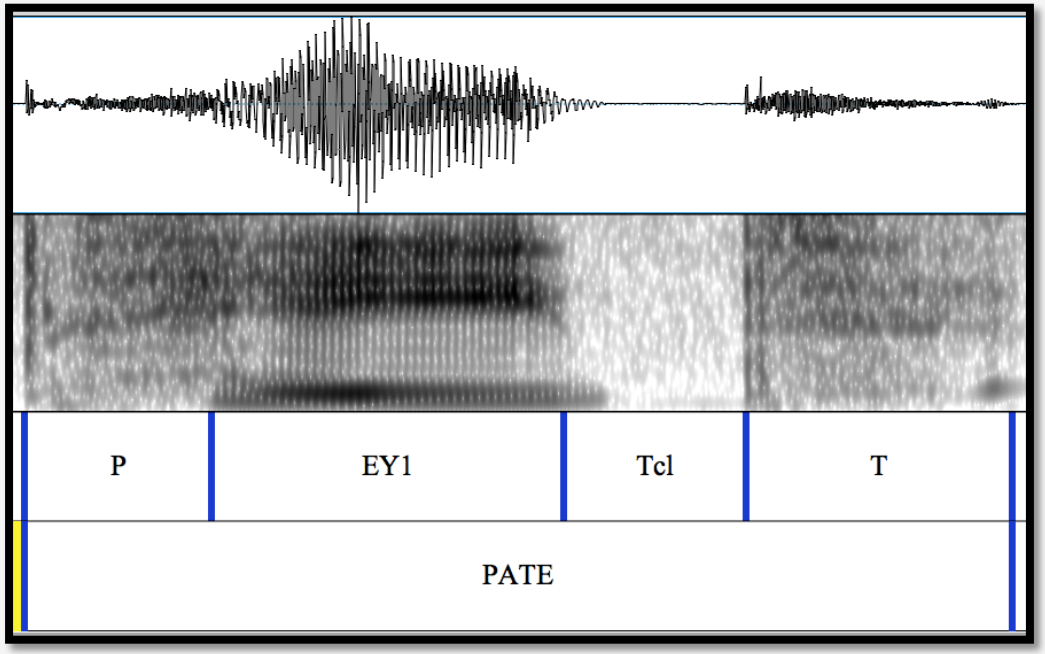
Background: Production

production study

laboratory and TIMIT experiments

methods adapted from Forrest et al. (1988), Jongman et al. (2000), Sundara (2005)

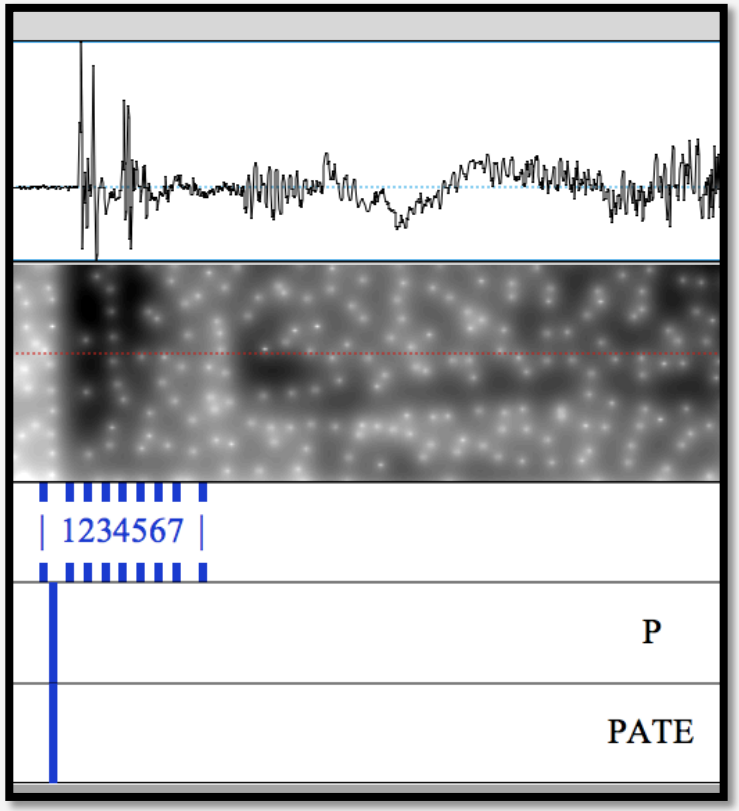
/p,t,k,b,d,g/ x /i,ɪ,e,ɛ,æ,ʌ,ɑ,ɔ,o,u/ x /t/



N=18 (4 male)
resampled at 16kHz
pre-emphasized above 1000Hz
high-pass filtered at 200Hz
segmented from transient to voicing

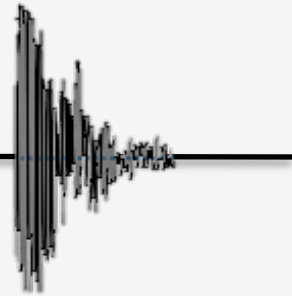
Laboratory Production: Methods

analysis as in Forrest et al. (1988), Hanson and Stevens (2003), Flemming (2007)

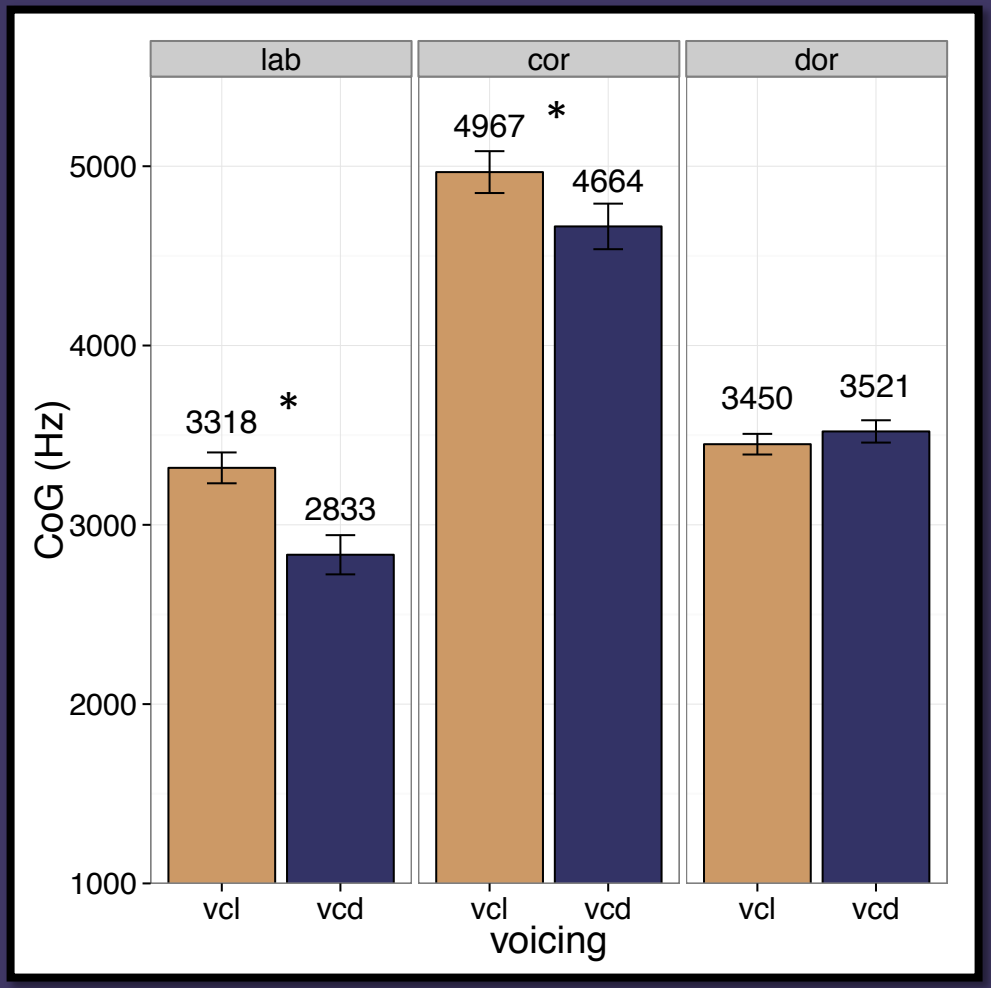


- Computed 64-point FFT for 7 consecutive 3ms Hamming windows, shifted by 1ms
- 7 PSDs averaged to give a smoothed spectrum
- Center of Gravity (CoG) calculated from smoothed spectrum: amplitude-weighted mean frequency

$$CoG = f_1 p(1) + \ldots + f_{32} p(32)$$



Laboratory Production: Measurement



Laboratory Production: Results

Mixed-effects linear regression

Fixed effects sum-coded and maximal random effect structure

voice $\beta_{voice} = 122$, $p < .01$

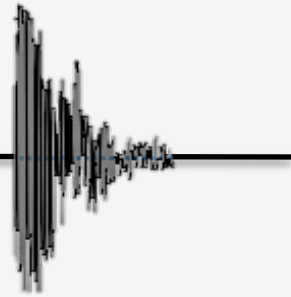
× place $\beta_{labial} = -633$, $p < .001$; $\beta_{coronal} = 916$, $p < .001$

× gender $\beta_{gender} = 86$, $p < .01$

Significant interactions examined with post-hoc comparisons

	labial	coronal	dorsal
male	$\beta_{voice} = 224$ $p < .001$	$\beta_{voice} = 224$ $p < .05$	n.s.
female	$\beta_{voice} = 253$ $p < .001$	n.s.	n.s.

Crucially, the pattern of significance remains the same when tokens with glottal pulses near the release are excluded.



Laboratory Production: Analysis

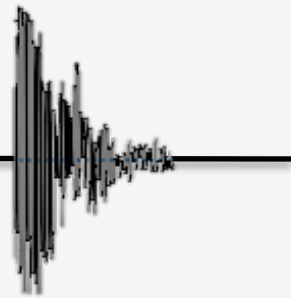
Byrd (1993), Keating et al. (1993)

630 different AE speakers

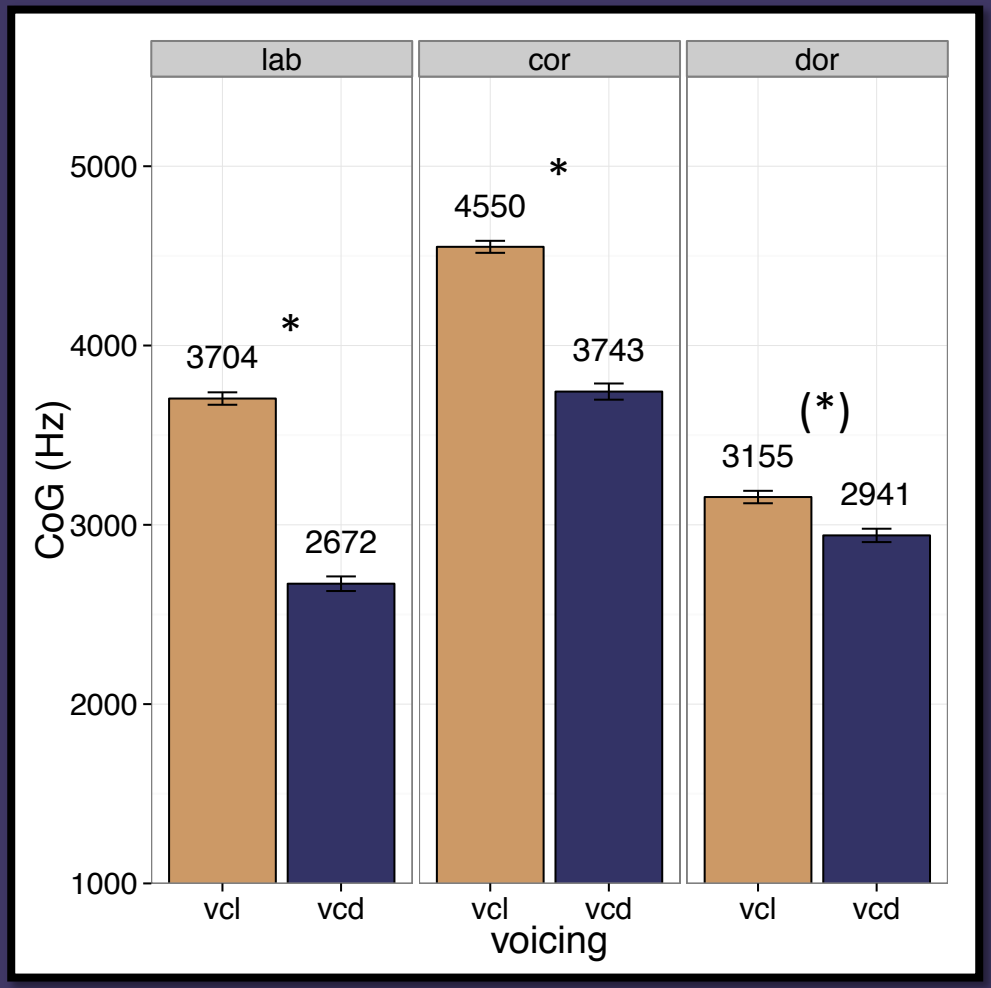
Word-initial, pre-vocalic /p, t, k, b, d, g/

Words with high token freq. removed (*too, to, do, carry, dark*)

Phoneme	Tokens	Phoneme	Tokens
/p/	661	/b/	668
/t/	579	/d/	547
/k/	1179	/g/	415



TIMIT: Methods



TIMIT: Results

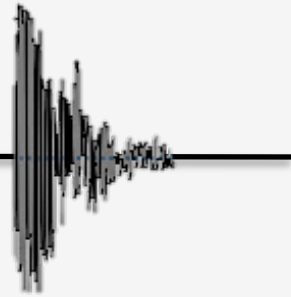
Mixed-effects linear regression
Fixed effects sum-coded and maximal random effect structure

voice $\beta_{voice} = 320$, $p < .001$
× place $\beta_{labial} = -314$, $p < .001$; $\beta_{coronal} = 762$, $p < .001$
× gender $\beta_{gender} = 205$, $p < .001$

Significant interactions examined with post-hoc comparisons

	labial	coronal	dorsal
male	$\beta_{voice} = 555$ $p < .001$	$\beta_{voice} = 460$ $p < .001$	($\beta_{voice} = 112$ $p < .001$)
female	$\beta_{voice} = 396$ $p < .001$	$\beta_{voice} = 280$ $p < .001$	($\beta_{voice} = 113$ $p < .05$)

Crucially, the pattern of significance remains the same, except for the dorsals, when tokens with glottal pulses near the release are excluded.



TIMIT: Analysis

perception study

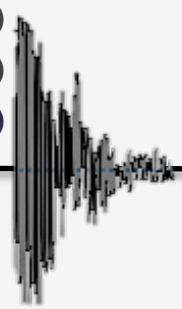
laboratory and Mechanical Turk experiments

/t/-burst VOT continuum

/d/-burst VOT continuum

Trading relation between burst and VOT

Keating (1979)
Nittrouer (1999)
Caldwell and Nittrouer (2013)

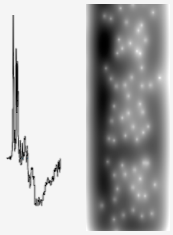


Background: Perception

Keating (1979), Ganong (1980), Andruski et al. (1994)

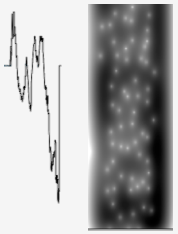
Labial Continua /bæt/-/pæt/

p

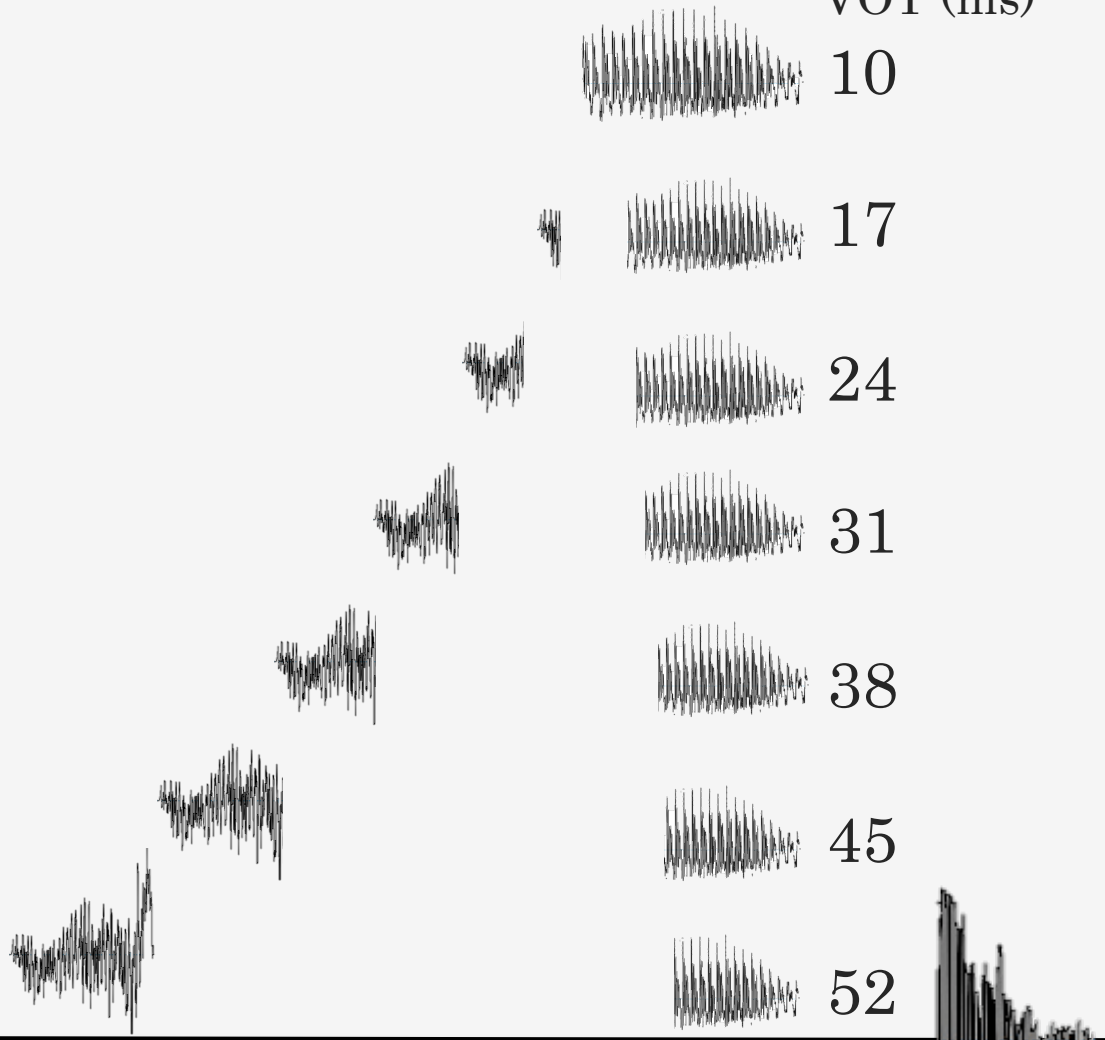


CoG: 3494Hz
Dur: 10ms

b



CoG: 1513Hz
Dur: 10ms

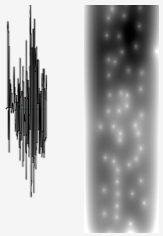


Laboratory Perception: Stimuli

Keating (1979), Ganong (1980), Andruski et al. (1994)

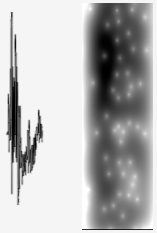
Coronal Continua /dat/-/tat/

t

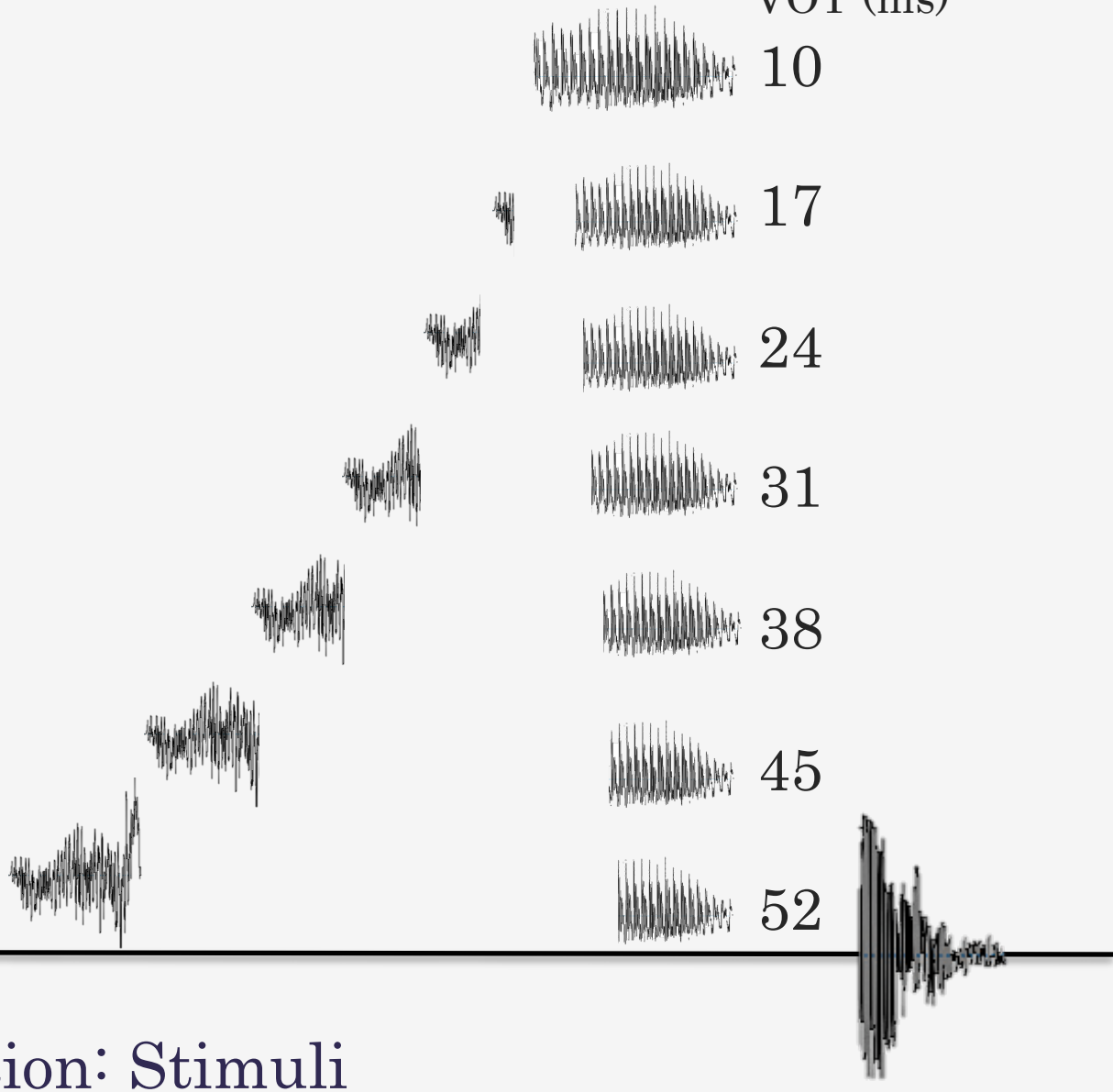


CoG: 5424Hz
Dur: 10ms

d



CoG: 3601Hz
Dur: 10ms



Laboratory Perception: Stimuli

Massaro and Cohen (1983), Hallé and Best (2007)

Two-alternative forced choice identification

T D

Differences verified with logistic mixed-effects analysis with maximal random effect structures

Goodness rating

T

very poor ok excellent

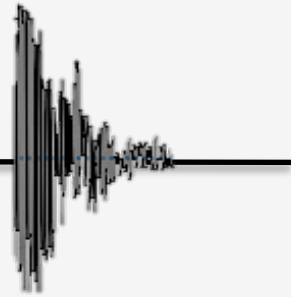
1 7

4

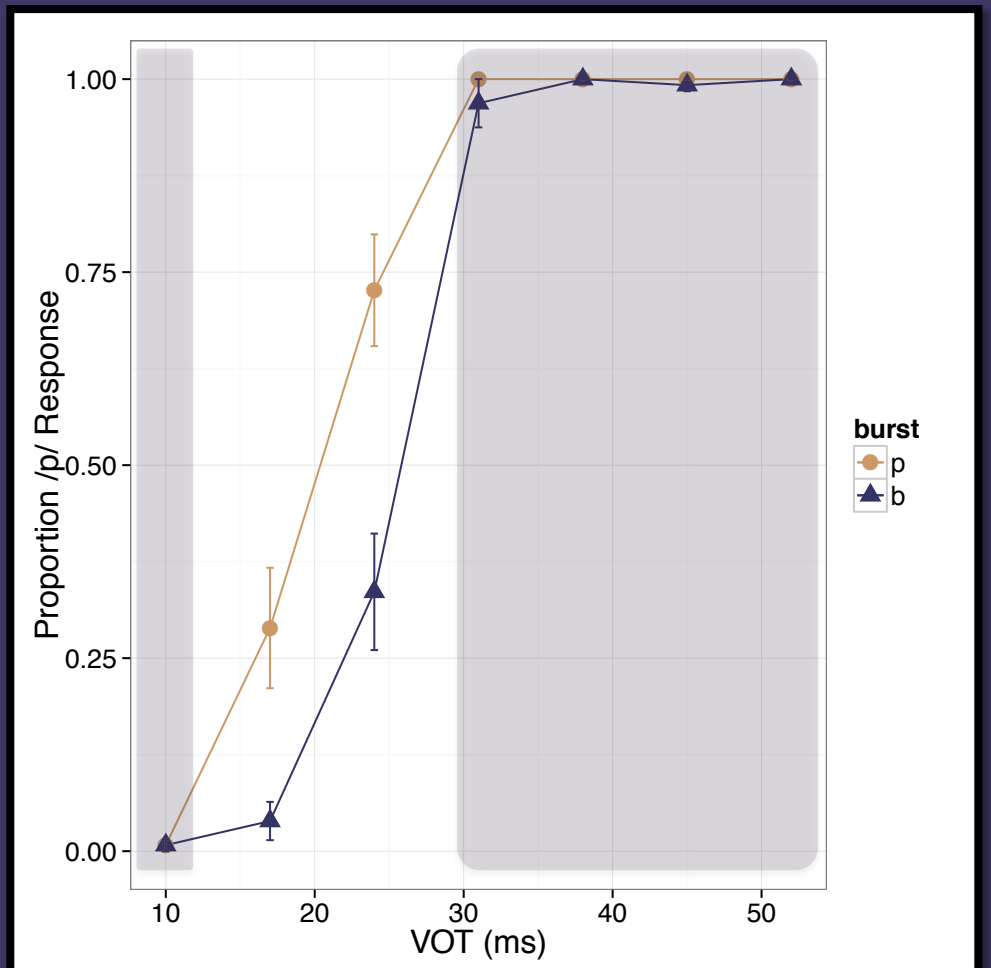
Differences verified with linear mixed-effects analysis with maximal random effect structures

Order of labial and coronal conditions counterbalanced

Within condition: 8 blocks of 14 stimuli in random order



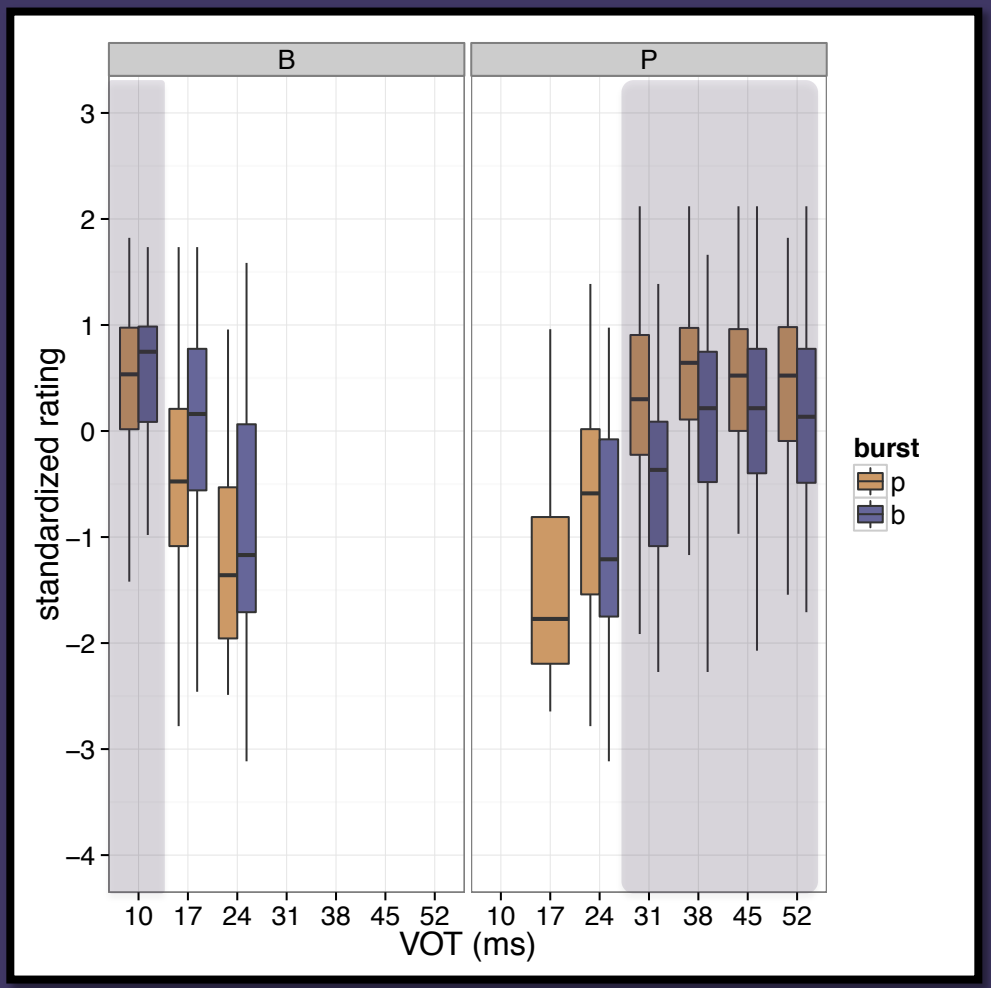
Laboratory Perception: Methods and analysis



labials

β_{burst} = .54
p<.001
N=16

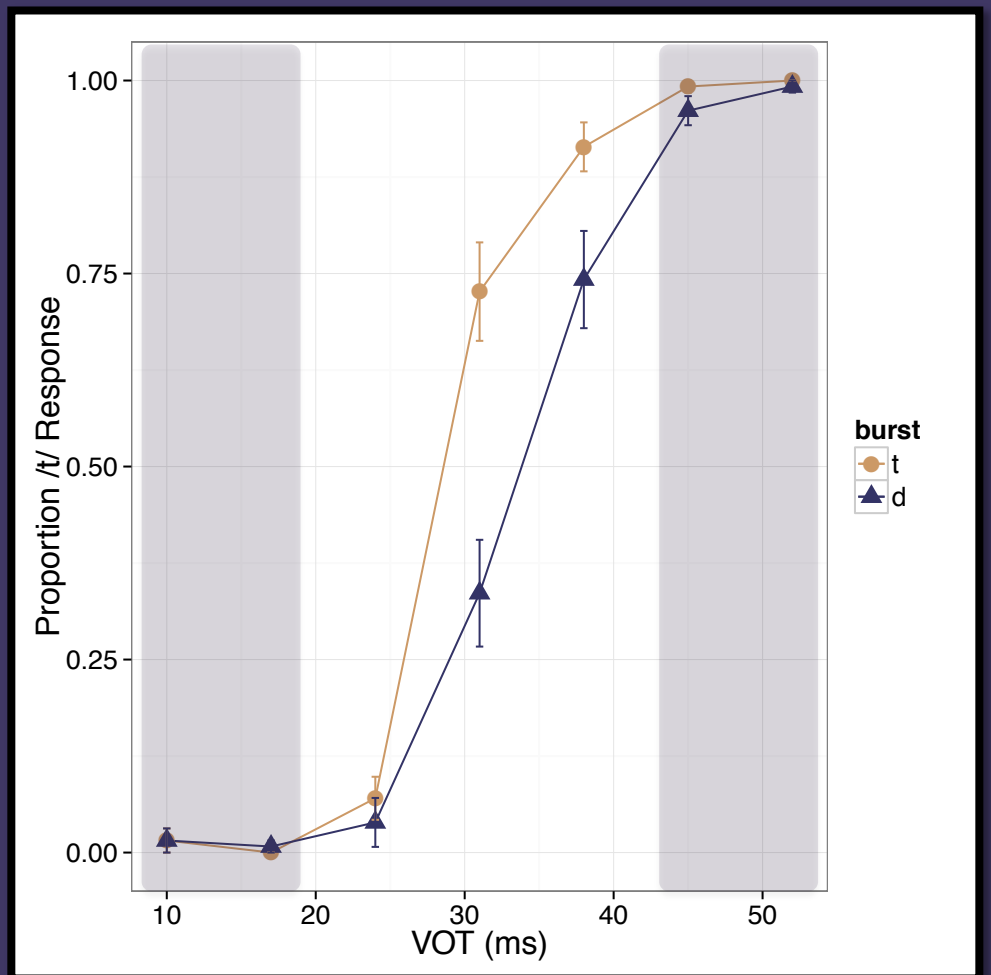
Laboratory Perception: Results



labials

N=16

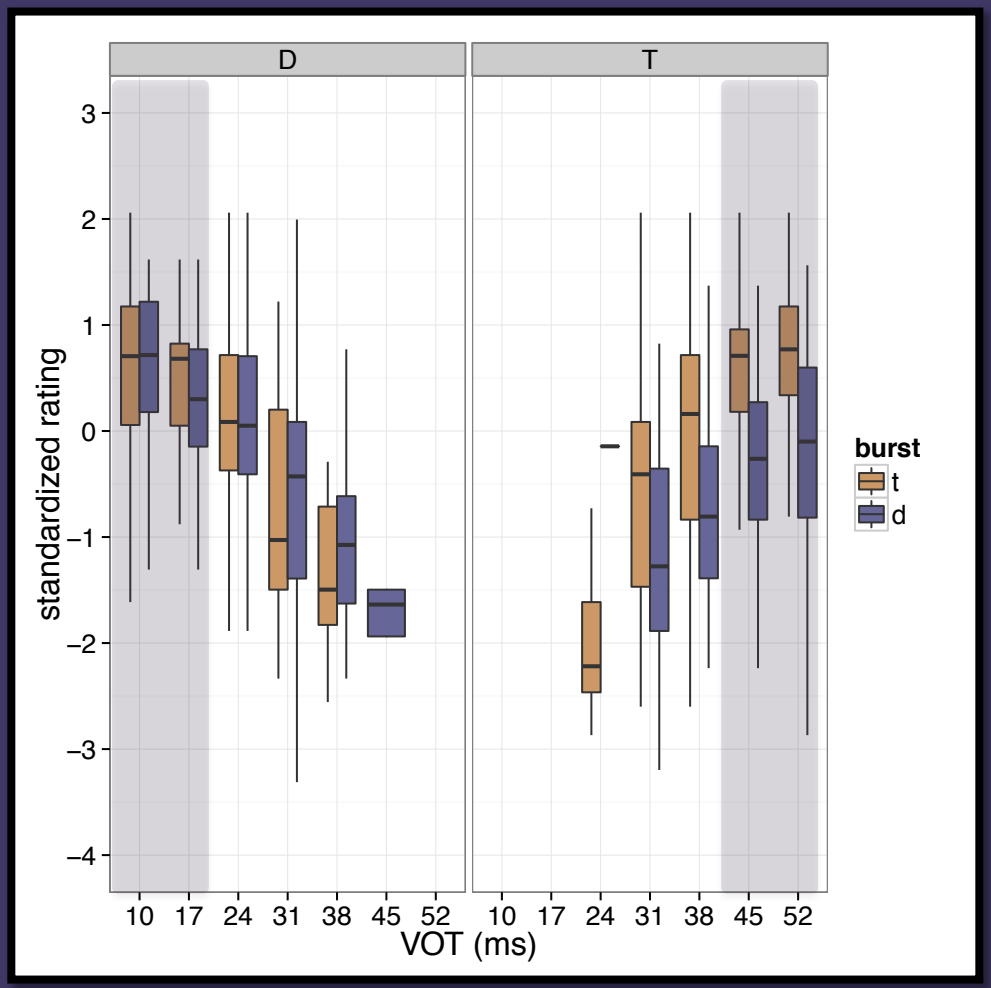
Laboratory Perception: Results



coronals

β_{burst} = .85
p<.001
N=16

Laboratory Perception: Results



coronals

N=16

Laboratory Perception: Results

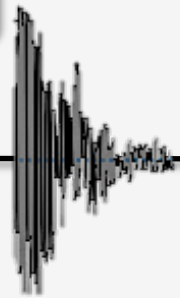
Kleinschmidt and Jaeger (2012), Eskanazi et al. (2013)

Crowdsourcing service increasingly used in psycholinguistics and phonetic studies

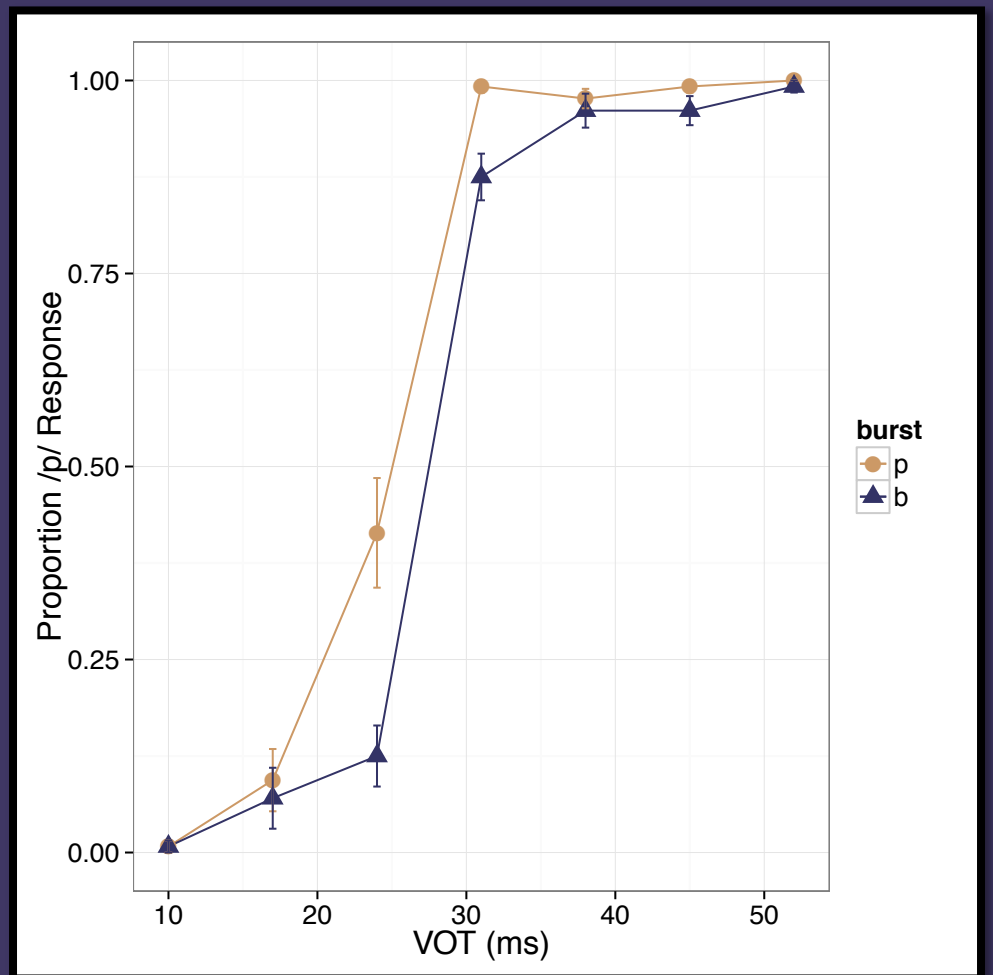
Greater diversity in participant population and listening conditions (noise!)

Labials
- 12 headphones
- 3 external speakers
- 1 internal speakers

Coronals
- 9 headphones
- 4 external speakers
- 3 internal speakers



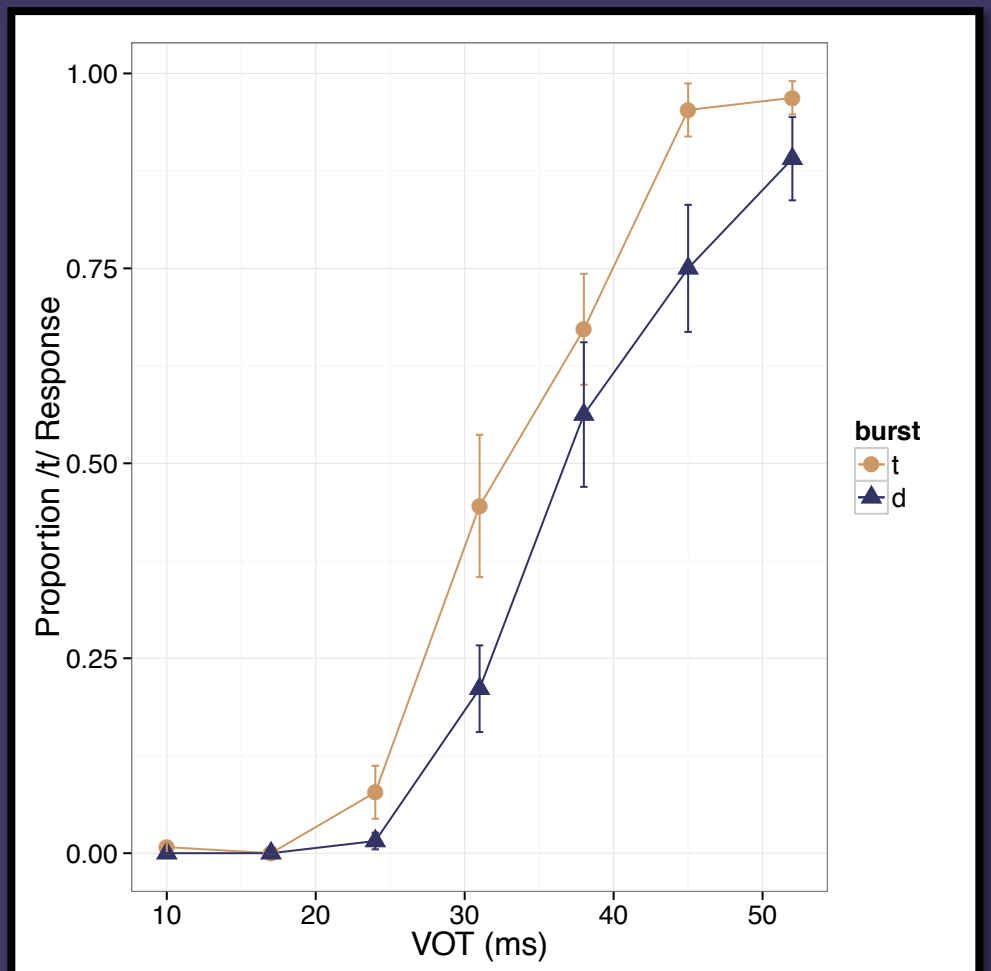
Mechanical Turk: Methods



labials

β_{burst} = .46
p<.001
N=16

Mechanical Turk: Results



coronals

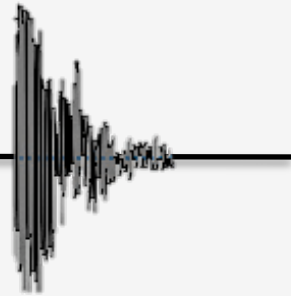
β_{burst} = .60
p<.001
N=16

Mechanical Turk: Results

Spectral shape of the burst is a cue to anterior stop consonant voicing

Higher CoG for voiceless labials and coronals

Spectral shape influences voicing identification



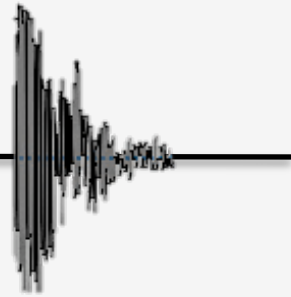
Summary and Implications

Repp (1978), Allopenna et al. (1998), Benkí (2001),
Stevens (2002), McMurray et al. (2008a)

Place and voice perception are interdependent

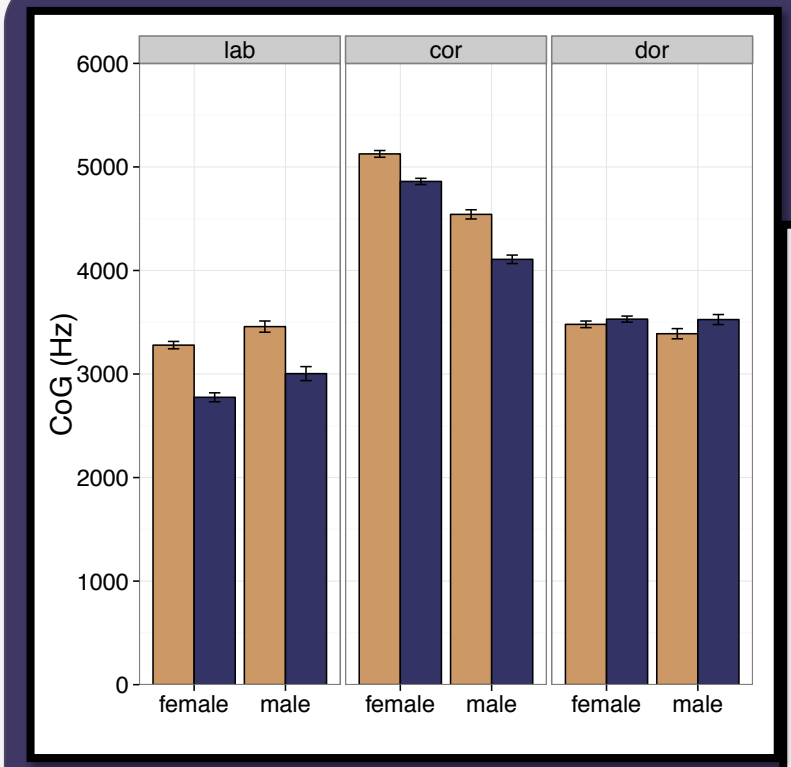
Cues to phonetic distinctions at burst landmark

Early cue to voicing and incremental perception



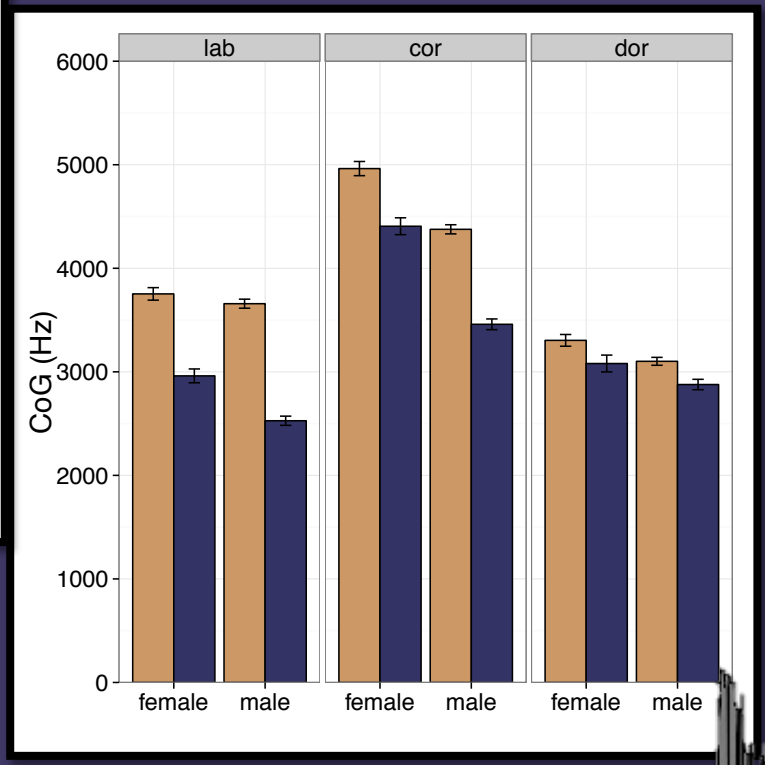
Summary and Implications

Thank you!



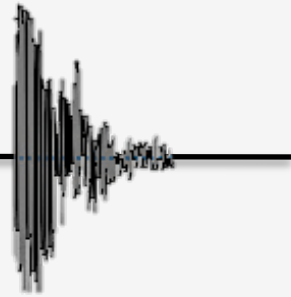
laboratory

TIMIT

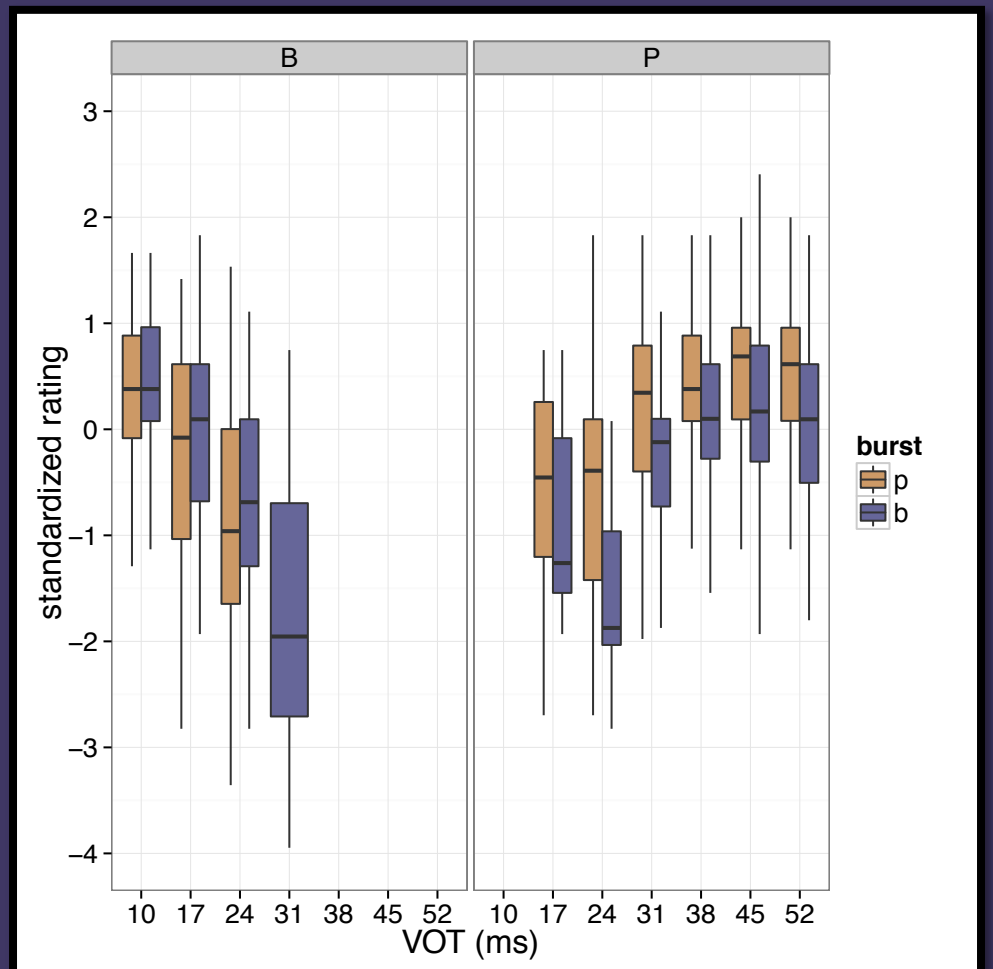


Production: Results by Gender

	Labial		Coronal		Dorsal	
Vowel	/p/	/b/	/t/	/d/	/k/	/g/
/i/	3582	3473	5278	5024	4031	4108
/ɪ/	3452	3107	--	4999	3893	3882
/e/	3447	2822	5253	4849	3868	3884
/ɛ/	3196	2587	5192	4777	3796	3840
/æ/	3142	2507	5161	4627	3725	3816
/ʌ/	3329	2694	5079	4682	3190	3291
/u/	3597	3247	4734	4534	3270	3172
/o/	3250	2830	4635	4400	3115	3170
/ɔ/	3150	2508	4697	4244	2713	2868
/ɑ/	3123	2575	4712	4460	2953	3254



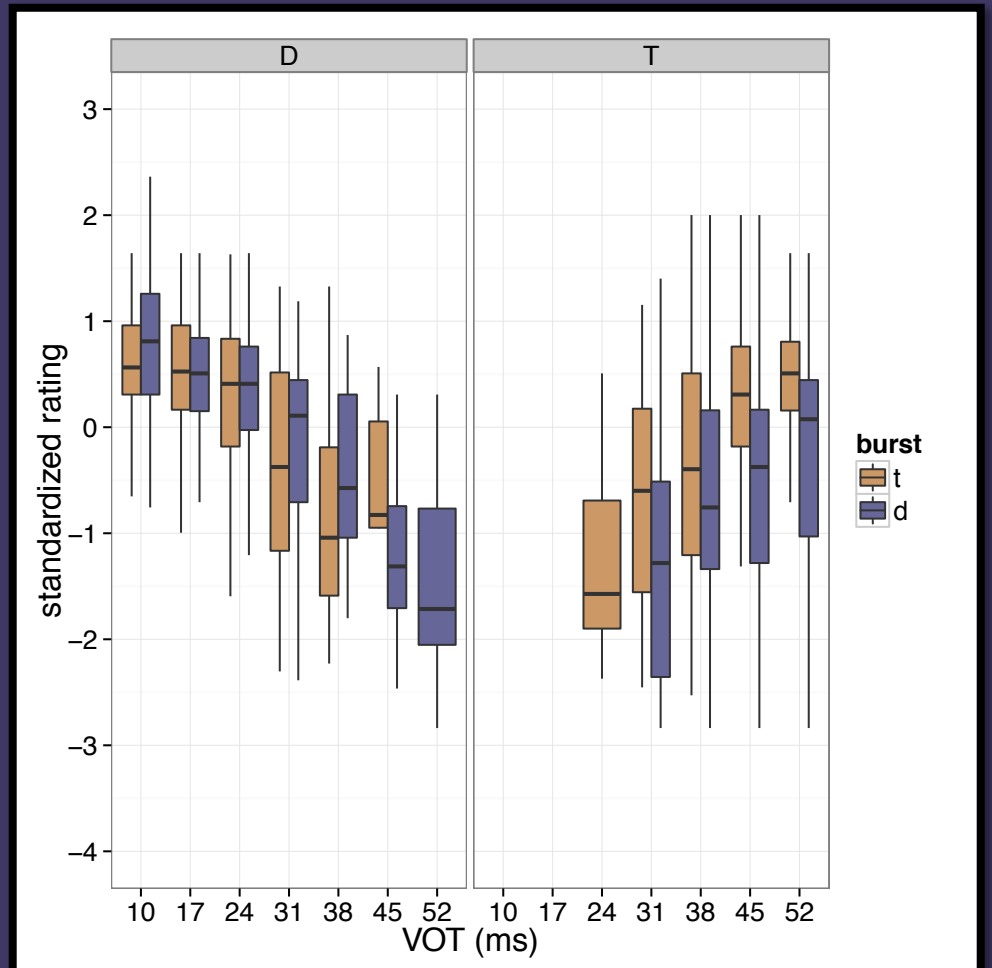
Production: Results by Vowel



labials

N=16

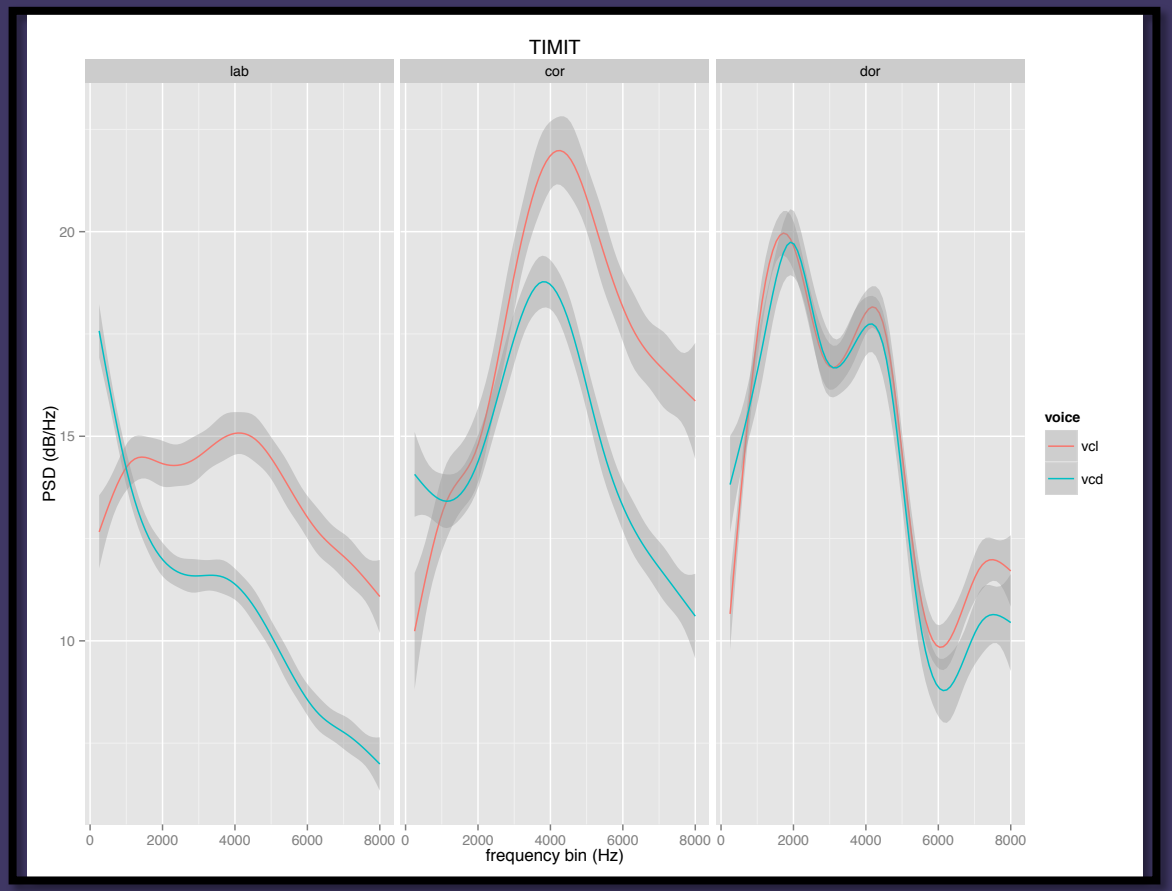
Mechanical Turk: Results



coronals

N=16

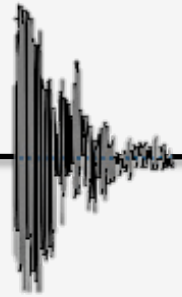
Mechanical Turk: Results



TIMIT: Spectral Shape

Study	Language	Measure	/p/	/b/	/t/	/d/	/k/	/g/
Zue 1976	Am. English	Peak	--	--	3600	3300	1940	1910
Parikh and Loizou 2005	Am. English	Peak	1910	1163	5649	5225	2261	2268
Sundara 2005	Ca. English	CoG	--	--	4900	4400	--	--
Kirkham 2011	Br. English	CoG	--	--	5220	4888	--	--
Van Alphen and Smits 2004	Dutch	CoG	1160	830	3540	2140	--	--
Sundara 2005	Ca. French	CoG	--	--	3800	3000	--	--
Vicenik 2010	Georgian	CoG	4000	3200	5300	4600	3100	3100

CoG = Center of Gravity (mean frequency)



Background: Production