Medial stop reduction and word recognition “in the wild”

Benjamin V. Tucker*, Daniel Brenner**, Michelle Sims*

*Department of Linguistics, University of Alberta
**Department of Linguistics, University of Arizona

Mental Lexicon, October 1, 2014

Acknowledgements

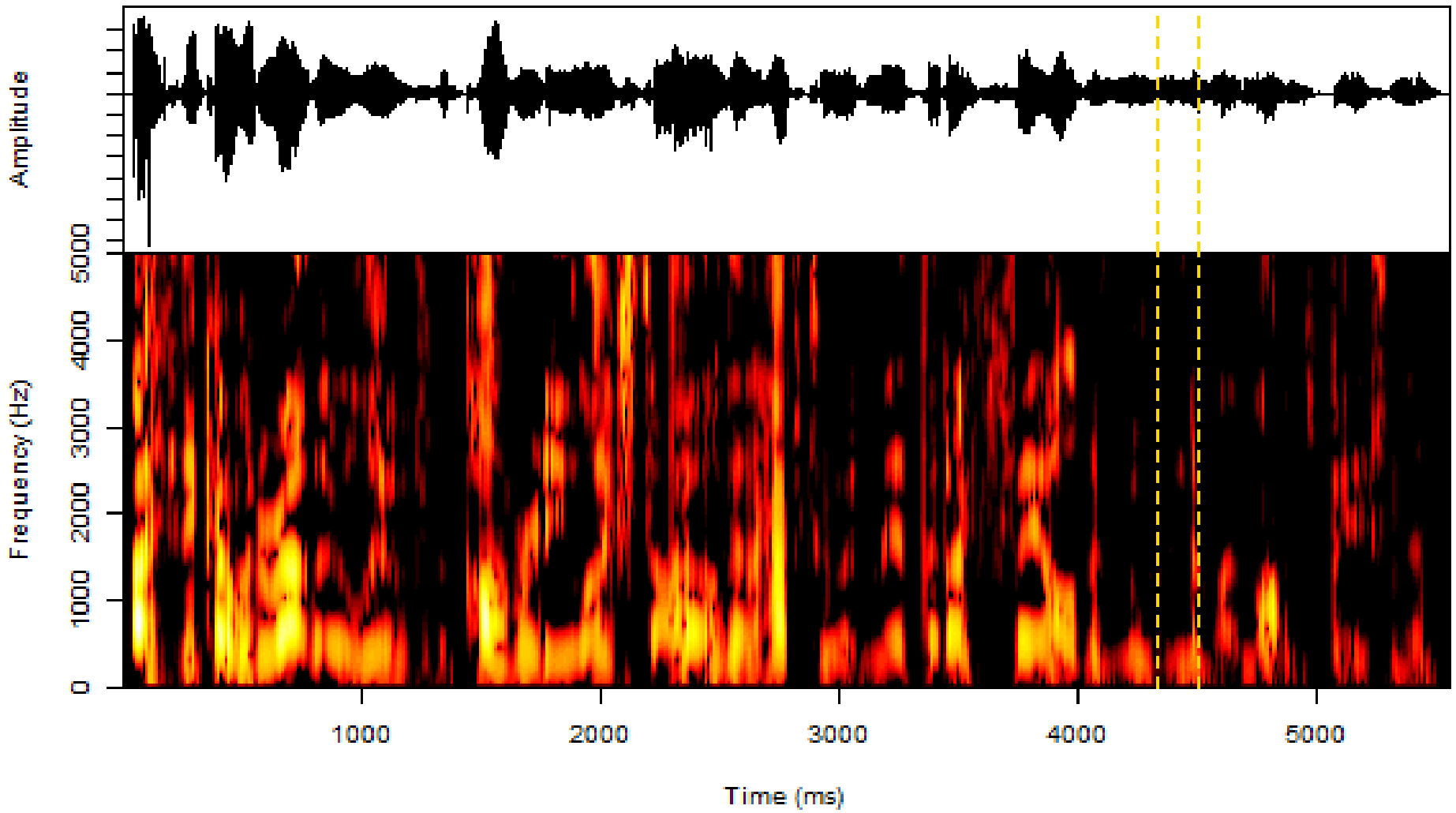
- Thanks to Janelle Dickout
- With funding provided by:

Conseil de recherches en sciences humaines du Canada

Social Sciences and Humanities Research Council of Canada

Canada

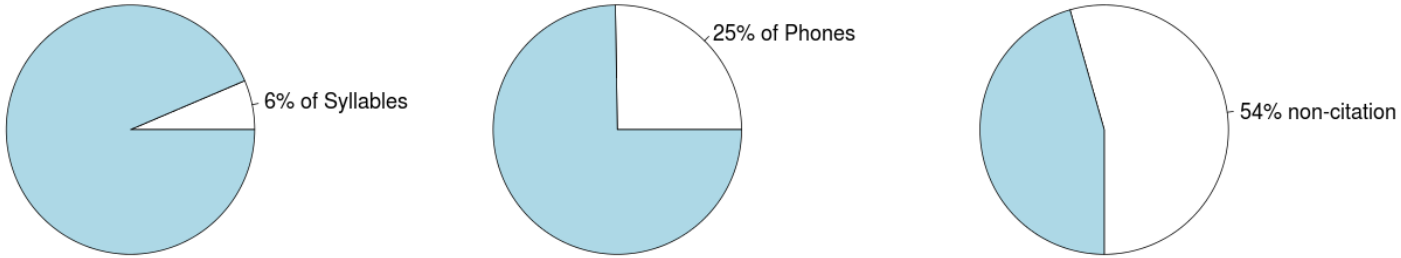
Why spontaneous speech?



target – full

Spontaneous and Conversational Speech

- Dilts (2013) shows using the Buckeye Corpus that 6% of syllables are deleted and 25% of segements are deleted.



- Greenberg (1999) shows using the Switchboard Corpus that:
  'that' is produced 117 different ways: [ðæ] is the most frequent at 11%
  'and' is produced 80 differenct ways: [æn], [ɛn], [ɨn], [ən], [n̩], [n], and only then [ænd].

Spontaneous and Conversational Speech

Reduced forms in other languages (Ernestus & Warner, 2011):

French	*c' était*	/setɛ/	[stɛ]	'was'
Finnish	*niinku*	/niːŋku/	[nik]	'like'
German	*wagen*	/vaːgən/	[vaːŋ]	'car'
Japanese	*nihongo*	/nihoŋgo/	[ĩj̃õːj̃]	'Japanese (language)'
Mandarin	*bu zhi dao*	/bu tʃɻ dao/	[bəɻao]	'don't know'
Swedish	*som alla*	/somalːa/	[smala]	'as all'
Korean	*saenggakpoda*	/sɛŋgakpoda/	[sampoda]	'than expected'

Spontaneous and Conversational Speech

- Models of the lexicon and linguistic categories
  What in speech variation is due to low-level phonetic processes, and what is due to the grammar?
- L2 acquisition
  L2 speakers struggle with reduced forms, and can't ask native speakers to explain. "Excuse me, but what's 'dyuatame'?"
- Natural Language Processing
  ASR systems perform much poorer on conversational recognition tasks than on careful speech (Ernestus & Warner, 2011).

“unreduced” vs. “reduced”

Unreduced

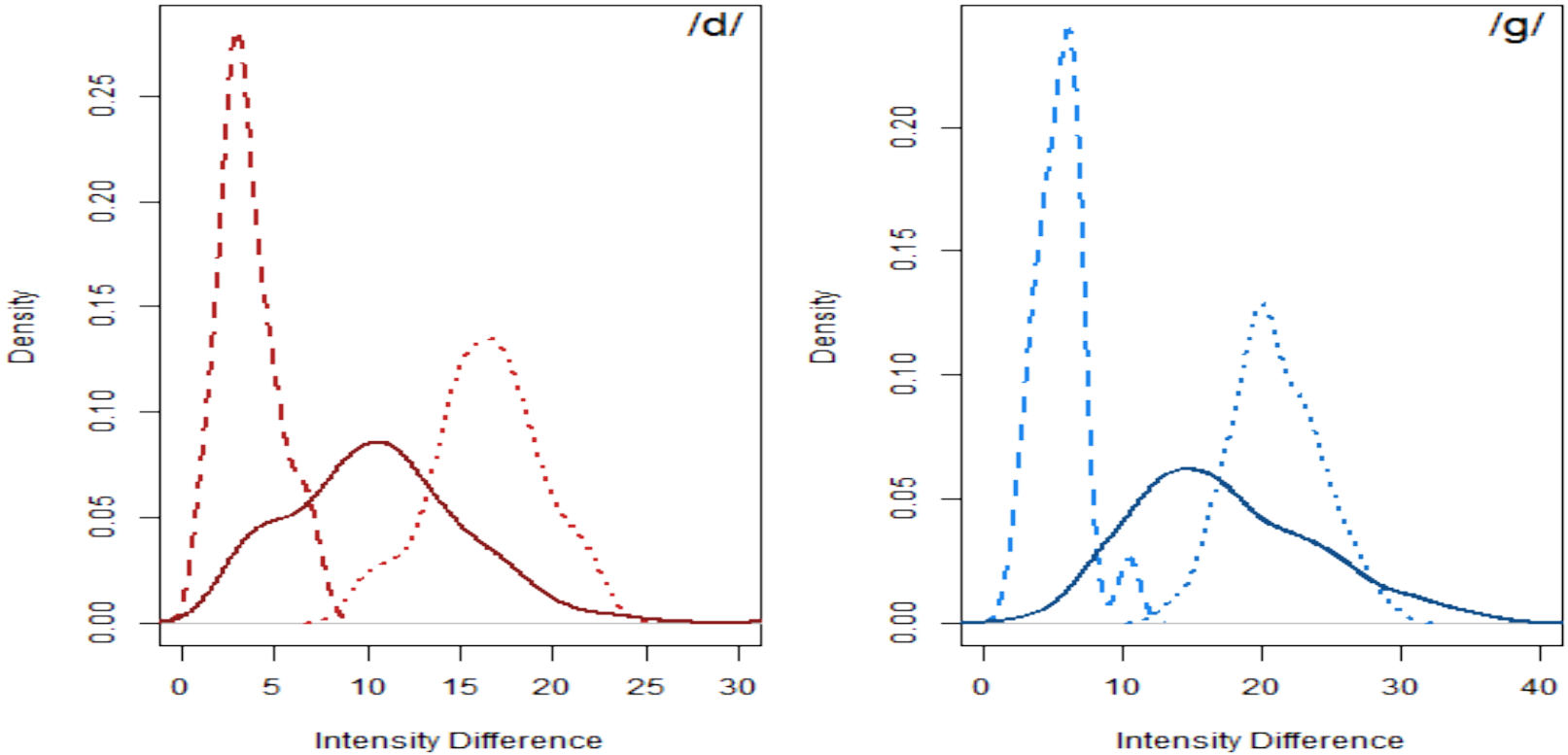
- Unreduced, “lab”, or careful speech has been the main genre of study in phonetics and psycholinguistics (Cutler, 1998; Warner, 2011).
- We know a lot about how words are produced and recognized in a “canonical” form.

Reduced

- Even amidst all this variation and uncertainty, listeners perform exceptionally well at spoken comprehension.

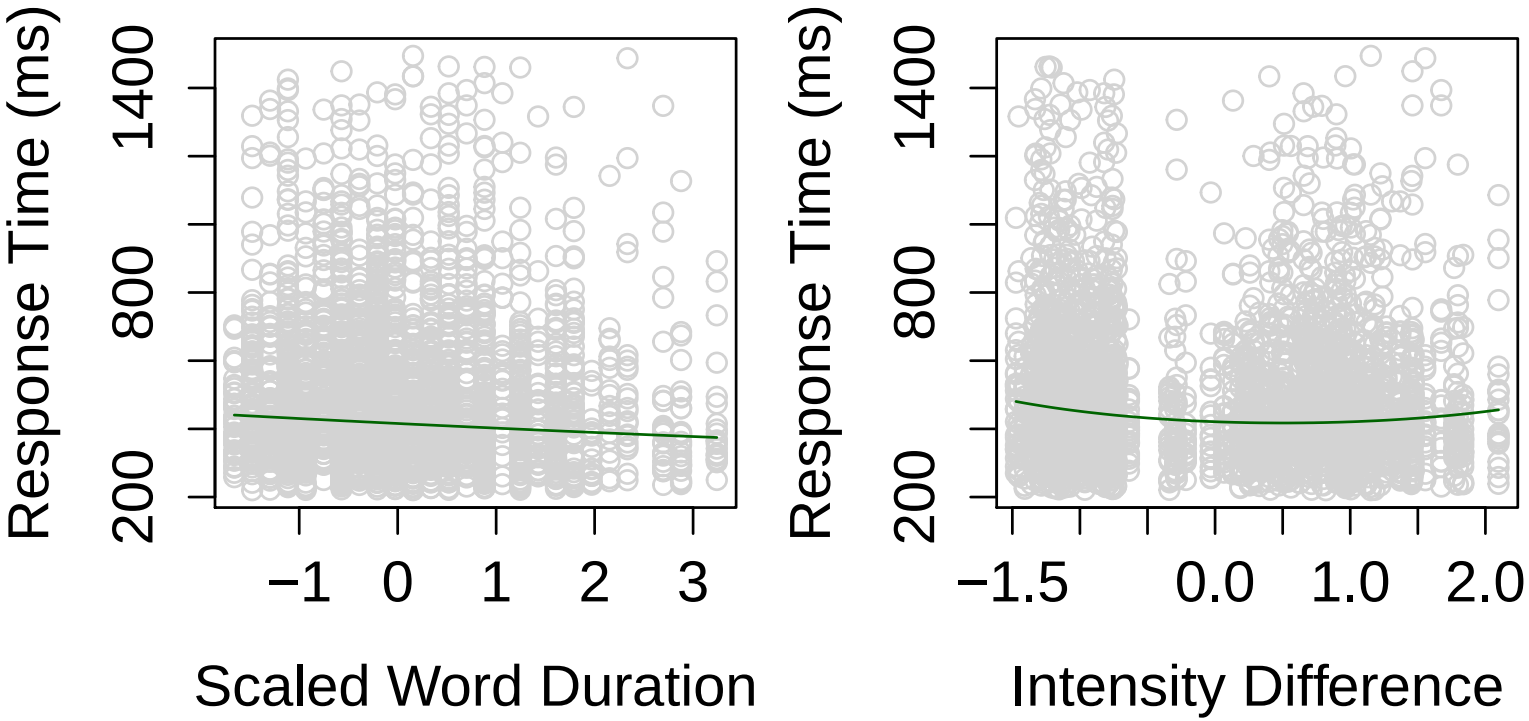
Word-medial stops: Production

(Tucker, 2011; Warner & Tucker, 2011)



Word-medial stops: Word Recognition

A post-hoc analysis: (Tucker, 2011)



Recognition in free context

What if we take our stimuli from a spontaneous conversation?

- Use items that are distributed across a range of possible variation.
  How do low level phonetic characteristics impact word recognition?
  What is the lexical representation of these sounds and words?

- What is the role of context in the recognition of casual word forms?

Recognition in free context

“The recognition of reduced word forms” (Ernestus, Baayen, & Schreuder, 2002)

- Degrees of reduction (high, med, low)
  mogelijk ‘possible’ [moxələk], [moxək], [mok]
- Contexts: Isolation, Phonological and Full
- Task: type what you hear

- They found that phonetic context and syntactic/semantic context increase identification accuracy regardless of the degree of reduction.

Current Study

- Adapt the design from (Ernestus et al., 2002)
- Use a speeded response task to tap into the aspects of lexical representaion

Method

Materials: 366 total items
All items were extracted from the telephone conversations reported in Warner and Tucker (2011)

- 56 target items
  - Degrees of reduction (measured by intensity)
  - Six types of word-medial stops [p/b, t/d, k/g]
  - Context: isolation, limited, and full
- Other items: Phonological overlap, Control, Fillers
- Auditory items were counterbalanced

Method

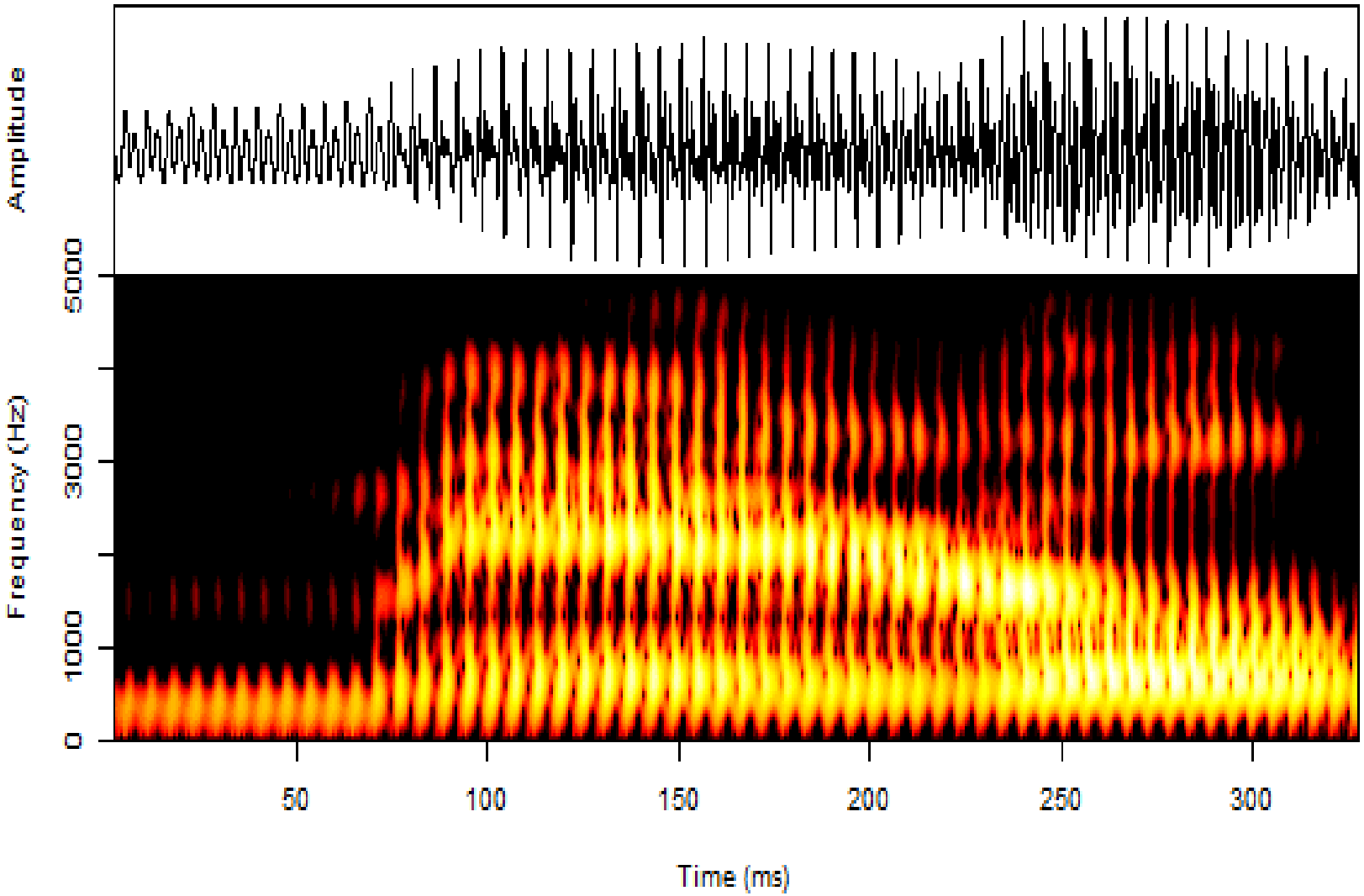
Subjects: 35 native speakers of Western Canadian English

- 28 females
- 7 males

Task: Cross-Modal Identity Priming

- Hear target in isolation or within its context (blocked)
- Printed word appears at onset of auditory prime
- Lexical decision on printed word (identity)
- Measure accuracy and reaction time

Method



target – limited – full

Analysis

Linear mixed-effects regression (Bates, Maechler, Bolker, & Walker, 2013)

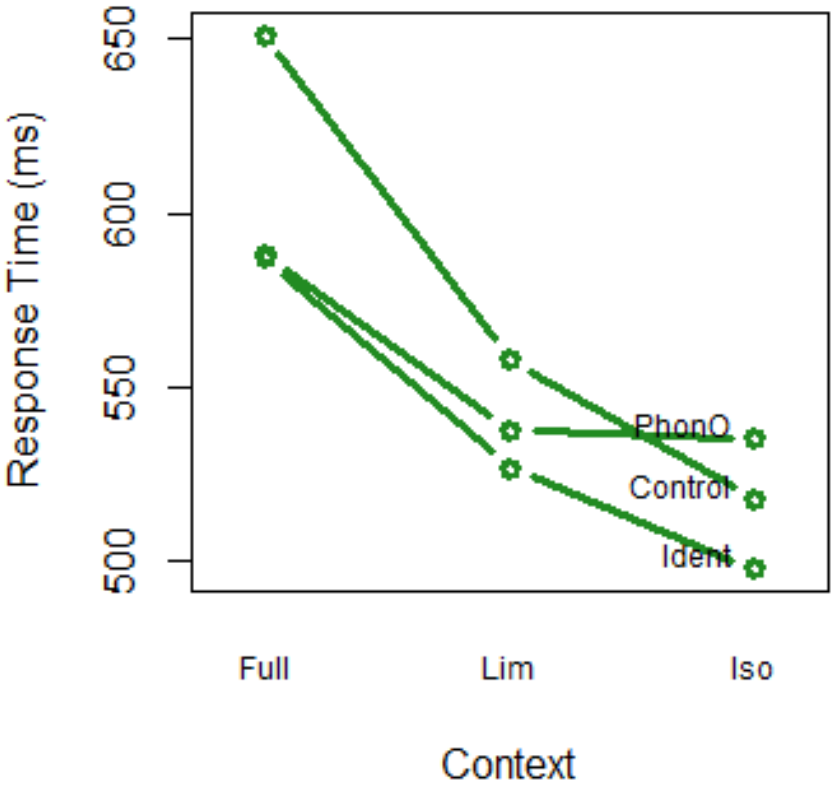
Predictors

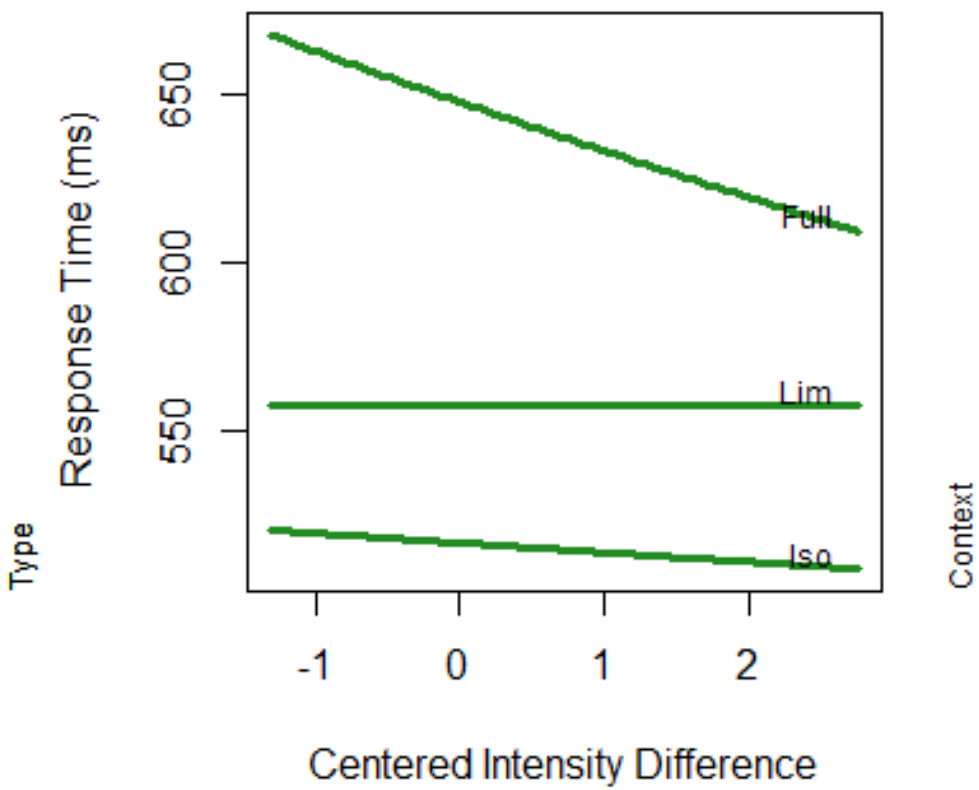
- Participant and Item (random effects)
- Intensity Difference
- Context (Full, Limited, Isolated)
- Word Duration (ms)
- Type (Phonological Overlap, Control, Identity)
- log Frequency

Results

- Significant priming effect
- Context and Type interact:
  - Control items get faster as the context decreases
  - Phonological Overlap items are faster than the control in all but the isolation condition
  - Identity items get faster as context decreases and is always faster than the Control items
- Intensity difference and Context interact:
  - Faster responses for less reduced word-medial stops in Full context and Isolation
  - No change for the Limited context stimuli

Results


Response Time (ms)
500
550
600
650
Full
Lim
Iso
Context
PhonO
Control
Ident


Response Time (ms)
550
600
650
-1
0
1
2
Centered Intensity Difference
Type
Context
Full
Lim
Iso

Discussion

- The context relationship differs from (Ernestus et al., 2002). Full context does not seem to facilitate faster responses.
- This is likely due to task differences, here the task relies on priming to facilitate response latency. The additional lexical and phonological information in the context compete with the target and slow down the response in this speeded lexical task.
- It would be beneficial to test the timing between the auditory prime and the visual target (identity priming can be finicky with this).
- The linear effect of degree of reduction is at odds with the distributional phoneme account of word recognition.

Conclusions

- No distributional effects on the processing of word-medial stops in this experiment
- In a speeded word recognition task with priming, additional context is not helpful in processing individual words
- We need more data!

Thank you!

References I

Bates, D., Maechler, M., Bolker, B., & Walker, S. (2013, October). *lme4: Linear mixed-effects models using eigen and s4.* Retrieved 2014-01-17, from `http://cran.r-project.org/web/packages/lme4/index.html`

Cutler, A. (1998). The recognition of spoken words with variable representations. In *Proceedings of ESCA workshop on sound patterns of spontaneous speech* (pp. 83–92).

Dilts, P. C. (2013). *Modelling phonetic reduction in a corpus of spoken english using random forests and mixed-effects regression* (Thesis). Retrieved 2014-02-03, from `https://era.library.ualberta.ca/public/view/item/uuid:383d9a26-35f7-43f9-8624-25ce5d2011ed`

References II

Ernestus, M., Baayen, R. H., & Schreuder, R. (2002). The recognition of reduced word forms. *Brain and Language*, *81*, 162–173.

Ernestus, M., & Warner, N. (2011). An introduction to reduced pronunciation variants. *Journal of Phonetics*, *39*, 253 - 260.

Greenberg, S. (1999). Speaking in shorthand - a syllable-centric perspective for understanding pronunciation variation. *Speech Communication*, *29*, 159–176.

Tucker, B. V. (2011). The effect of reduction on the processing of flaps and /g/ in isolated words. *Journal of Phonetics*, *39*(3), 312—318. Retrieved from `http://www.sciencedirect.com/science/article/pii/S0095447010000999` doi: 10.1016/j.wocn.2010.12.001

References III

Warner, N. (2011). Reduction. In M. van Oostendorp, C. Ewen, E. Hume, & K. Rice (Eds.), *The blackwell companion to phonology: General issues and segmental phonology* (Vol. 1, pp. 1866–1891). John Wiley & Sons.

Warner, N., & Tucker, B. V. (2011). Phonetic variability of stops and flaps in spontaneous and careful speech. *The Journal of the Acoustical Society of America*, *130*(3), 1606—1617. Retrieved 2011-12-21, from `http://asadl.org/jasa/resource/1/jasman/v130/i3/p1606_s1?isAuthorized=no` doi: 10.1121/1.3621306