The 14th Conference on Laboratory Phonology
NINJAL, Tachikawa, Japan
July 25-27, 2014

Acoustic variability aids the interpretation of phonetic detail in cross-language speech production

Lisa Davidson
Sean Martin
Colin Wilson

New York University
Johns Hopkins University

Learning within-category variability

- Languages differ in the permitted phonetic realization of sound structures
  - Classic case of VOT: Does 20-50ms lie within the category of voiceless aspirated stops (French) or span the boundary between two categories (Thai)? (Kessinger & Blumstein 1997)
- Interpreting fine-grained phonetic detail in L2 as ***within-structure variation*** or ***between-structure contrast*** is difficult when L1 differs (e.g., Iverson & Kuhl 1995; Best et al. 2001, 2003; Escudero & Boersma 2004, Escudero & Vasiliev 2011)

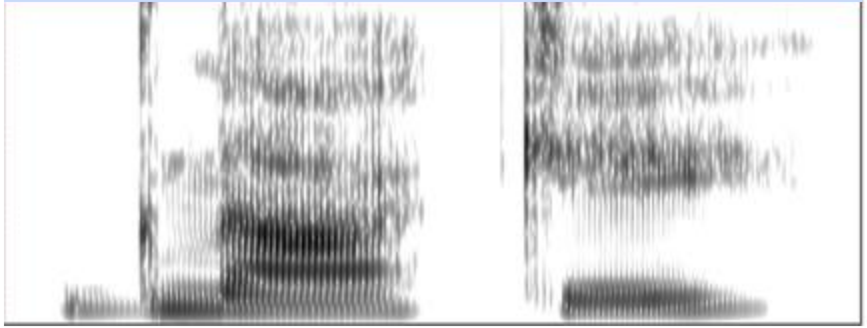
Learning within-category variability

- Exposure to diverse phonetic realizations appears to be crucial for acquiring the acceptable range of phonetic variation for each structure. (e.g., Lively, Logan, Pisoni 1993, Bradlow et al 1997, Wang et al 1999, Sommers and Barcroft 2007, 2011, Iverson et al. 2005, Seidl et al. 2013, Richtsmeier et al., 2009)

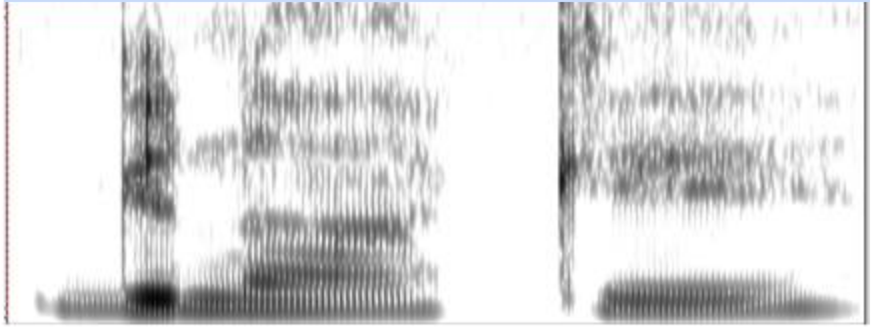
- A particularly effective type of phonetic diversity is provided by speech from multiple talkers. (Barcroft & Sommers 2005, Rost and McMurray 2010)

Interpreting phonetic detail in non-native phonotactics

Russian speaker (stimulus): [gnake]



English speaker reproduction: [g^{ə}nake]



When faced with non-native sequences, speakers often modify them to make them more like licit sequences (Davidson 2006, 2010): [g^{ə}nake] (epenthesis), [^{ə}gnake] (prothesis), [nake] (deletion), ...

Interpreting phonetic detail in non-native phonotactics

- Wilson and Davidson (2013) found evidence that categorical repairs of non-native sequences are due to ‘over-interpretation’ of fine-grained phonetic details such as voicing and burst duration.

- Post-hoc examination of production of initial clusters in Davidson (2010) identified several non-contrastive acoustic properties of Russian speech that are linked to particular English modifications.
  - E.g., more epenthesis in stop-initial clusters ([bdafe], [pkamo]) when the stop burst is longer or higher in amplitude

Current study

- Non-word Production Condition 1 (**Low Variability**)
  Do the productions of L2 speakers show sensitivity to non-contrastive phonetic detail in auditory stimuli?
  - Systematically manipulate identified acoustic properties of Russian stimuli beginning with C1C2 clusters that are illegal in English
- Non-word Production Condition 2 (**High Variability**)
  Does greater phonetic variability (tokens and talkers) reduce the influence of stimulus-specific phonetic detail?
  - Present the same manipulated stimuli from the Low Variability condition, but preceded by the same word spoken by 2 other talkers

Predictions

1. Low Variability condition: If sensitivity to phonetic realization strongly influences perception/production of a non-native target, acoustic manipulations should be mirrored by changes in performance
2. High Variability condition
    a. Selection: Speakers may select 1 token as a production model. If the last token is selected, modifications should pattern like the Low Variability condition.
    b. Abstraction: Speakers may ignore fine grained phonetic detail; responses should show little effect of acoustic manipulations (of last talker).
    c. Blending: Speakers productions may reflect a combination of all of the phonetic cues. Some effects of the manipulations may still be present, but they should be attenuated.

Stimuli

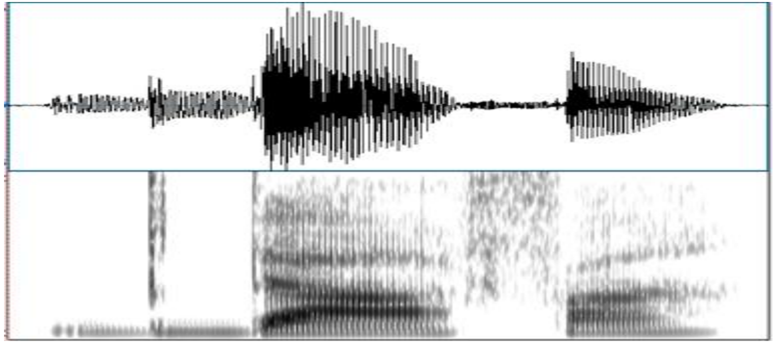
- All conditions: English speaking participants heard and repeated critical items of the form [C1C2áCV] produced by a native Russian speaker and corresponding fillers with initial and medial schwa
  - [ptake], [əptake], [pətake]; [zgamo], [əzgamo], [zəgamo]
  - Each cluster appeared in 4 distinct [___aCV] items

Cluster type	C1[-voice]	C1[+voice]
SS (stop-stop)	pt, tp, kp, kt	bd, db, gb, gd
SN (stop-nasal)	pn, tm, km, kn	bn, dm, gm, gn
FS (fricative-stop)	(not tested)	vd, vg, zb, zg
FN (fricative-nasal)	(not tested)	vm, vn, zm, zn

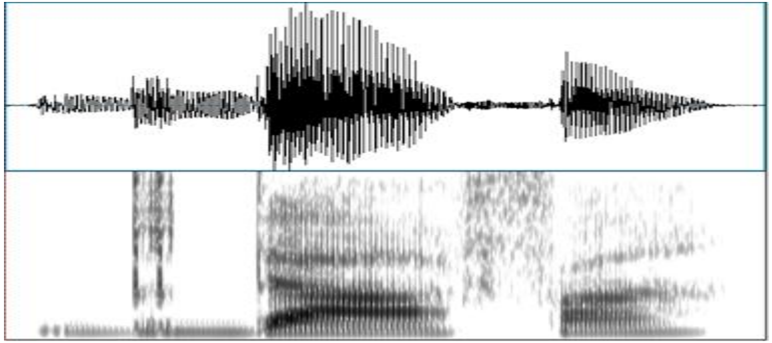
Stimuli: Acoustic manipulations

- **DUR**: duration of the acoustic transition (burst + aspiration) between stop and following consonant
  –2 levels: 20ms, 50ms ***–DUR longer → more epenthesis***
- **AMP**: amplitude of the acoustic transition of a stop (the burst) relative to the following consonant's amplitude
  –2 levels: high and low ***–AMP higher→ less deletion & C1 change, more epenthesis***
- **POV** (pre-obstruent voicing): interval of modal voicing *preceding* the onset of a voiced obstruent constriction
  –2 levels: present vs. absent ***–POV present → more prothesis***

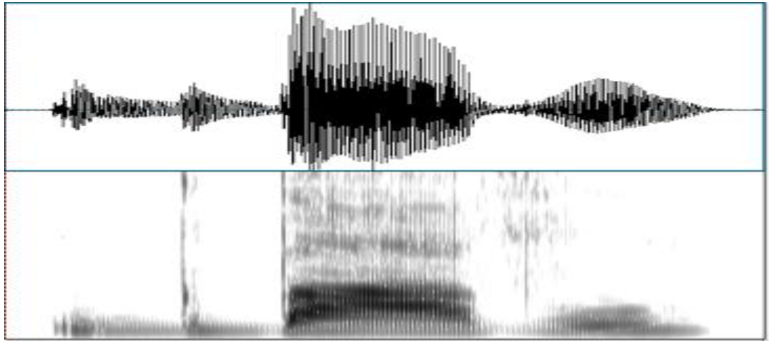
Stimuli: Examples



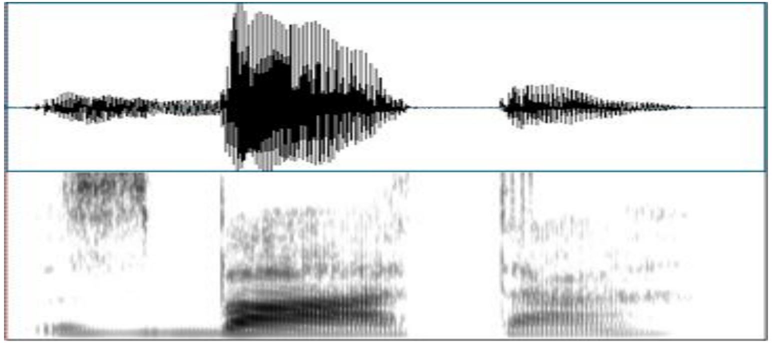
/bdafa/ 20ms burst



/bdafa/ 50ms burst



/gbavu/ with POV (20ms burst)



/zbata/ with POV

Procedure

Low Variability condition (Participants: N = 24 AmEng speakers)

- Identical stimulus item played twice then repeated by participant
- 32 each of FN, FS, SN, SS; 48 each of əCC and CəC fillers

High Variability condition (Participants: N = 24 AmEng speakers)

- Listeners heard each word produced 3 times:

  /gbavu/$_{\text{voice1}}$ /gbavu/$_{\text{voice2}}$ /gbavu/$_{\text{StimFromLowVariability}}$

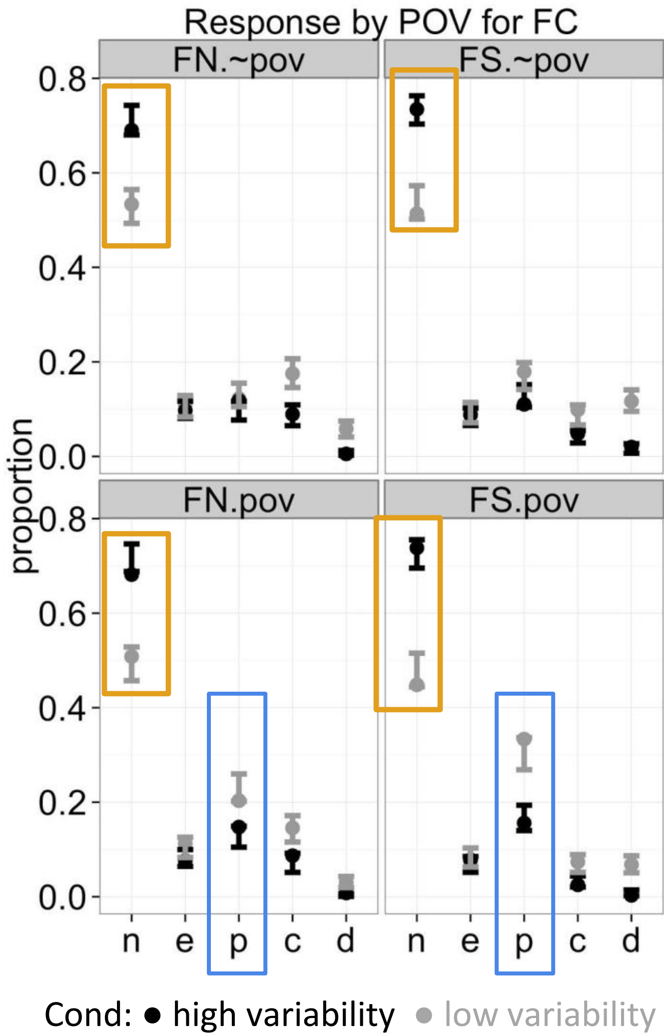
- Utterances for voice1 and voice2 were counterbalanced and set to baseline values (based on natural production averages)
  - 30ms DUR, high AMP for SS, low AMP for SN, no POV

Results: POV manipulation fricative-initial clusters

- No modification (correct) is the most typical response type.
- Prothesis is more prevalent when POV is present (bottom row) ($\beta = .68$, $p < .01$)
- Prothesis decreases in the high variability condition (black bars) ($\beta = -1.45$, $p < .01$)...
- ...but an effect of POV remains in the high variability condition ($\beta = .45$, $p < .01$)

All analyses are multinomial logistic regressions

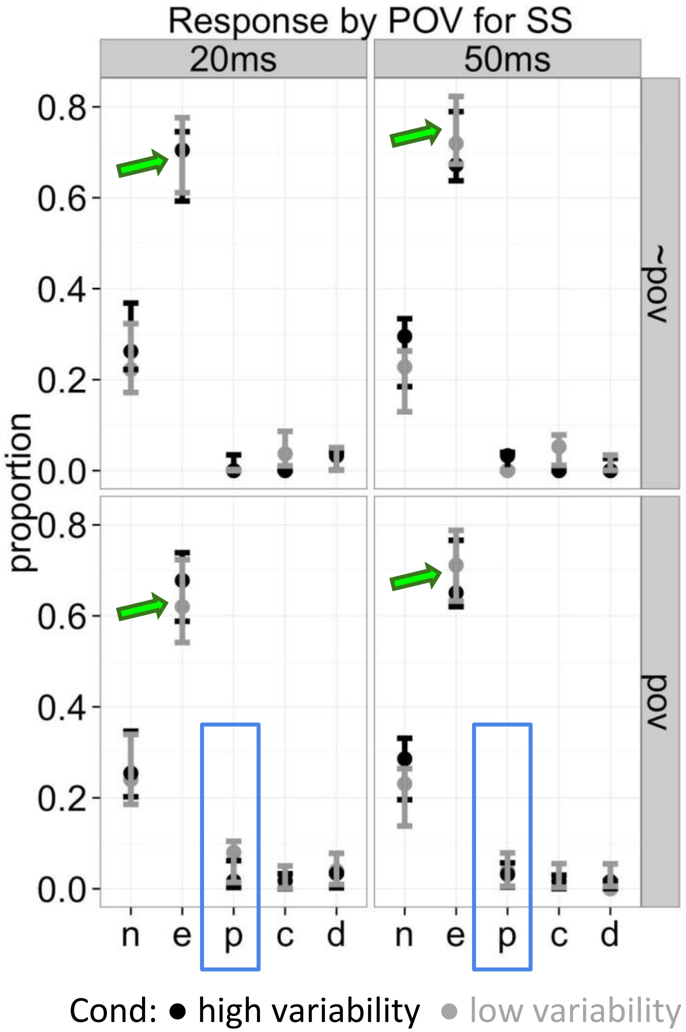
n = no modification
e = epenthesis
p = prothesis
c = C1 change
d = deletion



Results: POV manipulation stop-stop (voiced) clusters

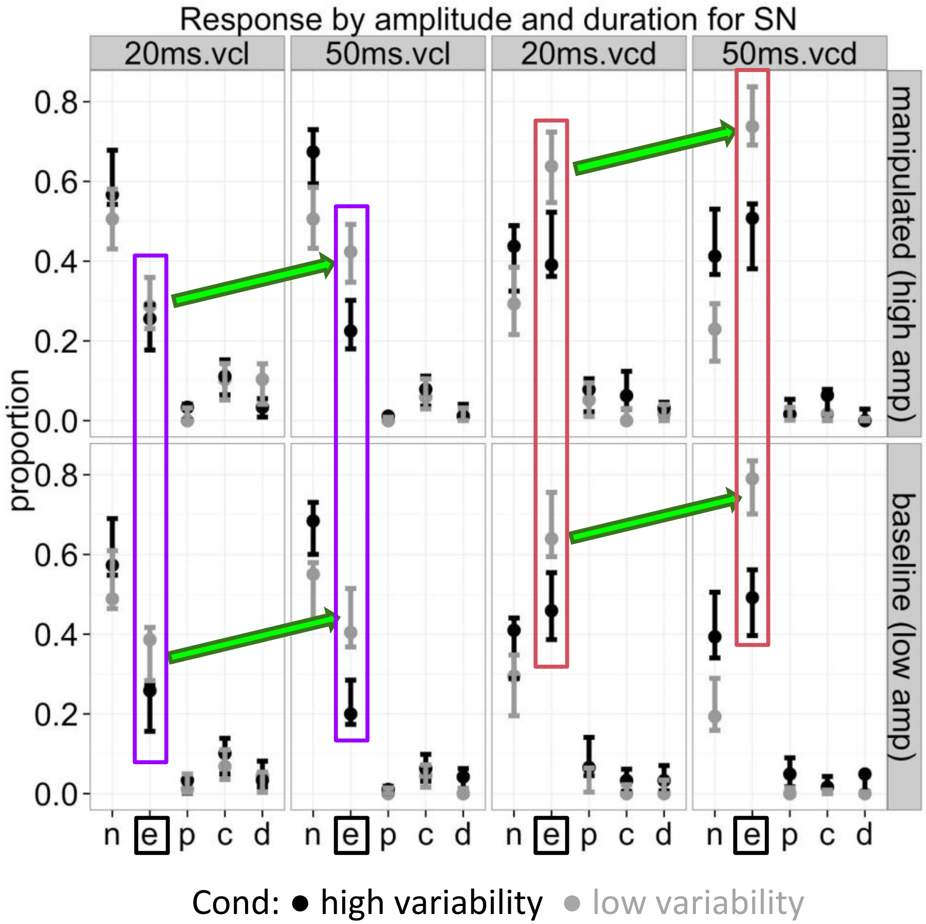
- No significant effect for SN clusters (not shown)
- Epenthesis was the most frequent response, but rate not affected by +POV
- Weak effect of +POV: significantly more prothesis for low variability condition only ($\beta = 1.12$, $p < .05$)

n = no mod
e = epenthesis
p = prothesis
c = C1 change
d = deletion



Results: DUR & AMP stop-nasal clusters

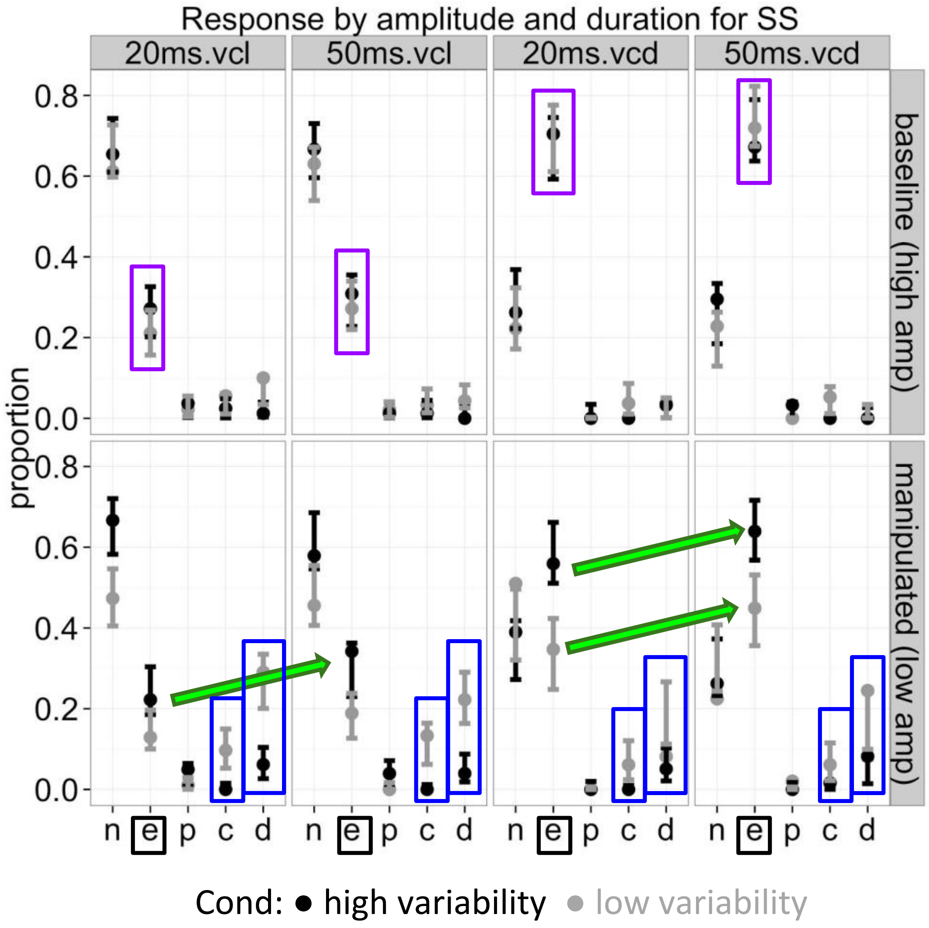
- **DUR**: Longer bursts increase likelihood of epenthesis in low variability (β = .34, p < .01) but not high variability condition. Less epenthesis in high variability overall (β = -1.15, p < .05)
- **AMP**: No effect of the amplitude manipulation either within or across variability conditions.
- Significantly less epenthesis for voiceless than voiced SN



n = no mod
e = epenthesis
p = prothesis
c = C1 change
d = deletion

Results: DUR & AMP stop-stop clusters

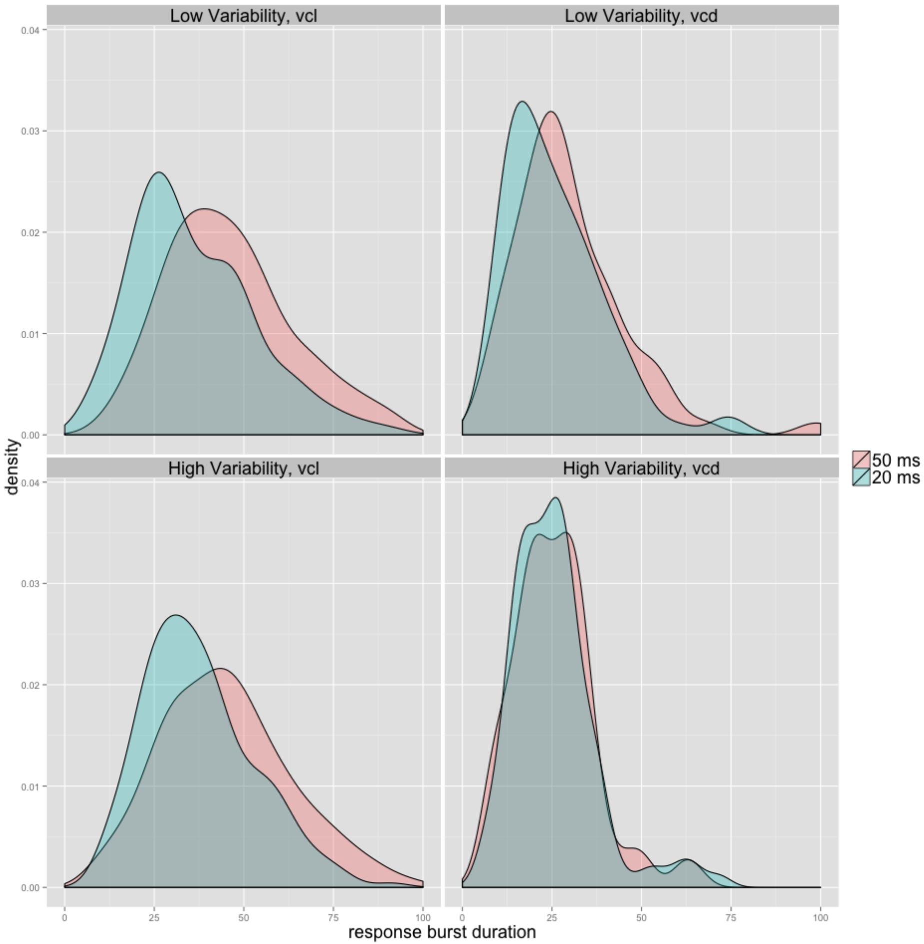
- **DUR**: Longer bursts marginally increase likelihood of epenthesis (β = .32, p = .07), but epenthesis is higher for *high variability* than low variability, because...
- **AMP**: Lower amplitude increases C1 change and deletion in the low variability condition. Higher amplitude increases epenthesis over all other response types for both conditions.
- Significantly less epenthesis for voiceless than voiced SS.



n = no mod
e = epenthesis
p = prothesis
c = C1 change
d = deletion

Results: Implementation of burst duration

- Are speakers sensitive to gradient phonetic detail? Is the effect attenuated by High Variability?
- Burst duration for no modification responses:
  - Significantly longer for 50ms ($\beta = 3.82$, $t = 7.60$)
  - Interaction between condition & duration: bursts are significantly longer in Low Variability for 50ms stimuli ($\beta = 1.61$, $t = 3.32$)



Summary of significant effects

Sequence	*Low Variability*	*High Variability*
FC	**POV**: More prothesis for +POV	**POV**: More prothesis for +POV (but less than low variability)
SN	**DUR**: More epenthesis for longer burst	***No effects***
SS	**DUR:** Marginally more epenthesis for longer burst **AMP:** More epenthesis for higher amp; More C1 del, change for lower amp **POV**: More prothesis with +POV	***No effects***

Overall production **more accurate** in high variability (56% correct) than low variability (39%) condition.

Phonetic influences on modifications

Especially in the low variability condition, manipulations have several effects:

- **Epenthesis**: English speakers interpret open transition after a stop as a cue to a vowel. Longer and louder bursts are more consistent with reduced vowels, as are voiced transitions.
- **Deletion and change**: For lower amplitude bursts, the release may be misperceived as a different stop, or not perceived at all.
- **Prothesis**: Strong voicing preceding and lasting throughout the obstruent closure may be perceived as a vowel, since English obstruents tend to be devoiced in initial contexts (e.g. Docherty 1992)

Effect of variability on the interpretation of phonetic detail

- Whereas the low variability condition demonstrated strong effects of the manipulations, these effects were greatly attenuated in the high variability condition.
  - But they did not disappear: +POV still led to more prothesis in the high variability case
  - Also, some imitation for increasing burst duration in both conditions, but *less* in high variability
- Most consistent with the blending account: speakers attend to and combine the phonetic information in all three utterances before producing the token.
- **L2 speakers can better identify target sound structures by blending (averaging) multiple variable productions of the same word.**

Thank you!

We would like to thank:

- Alice Hall, Francesca Himelman, Johnny Mkitarian, Elizabeth George, and Steven Foley for data analysis
- Members of the NYU PEP Lab & JHU Phonetics/Phonology Lab, & audiences at Penn, Delaware, Stony Brook, and the JHU IGERT Workshop for feedback
- National Science Foundation for grants BCS-1052855 to Davidson and BCS-1052784 to Wilson.