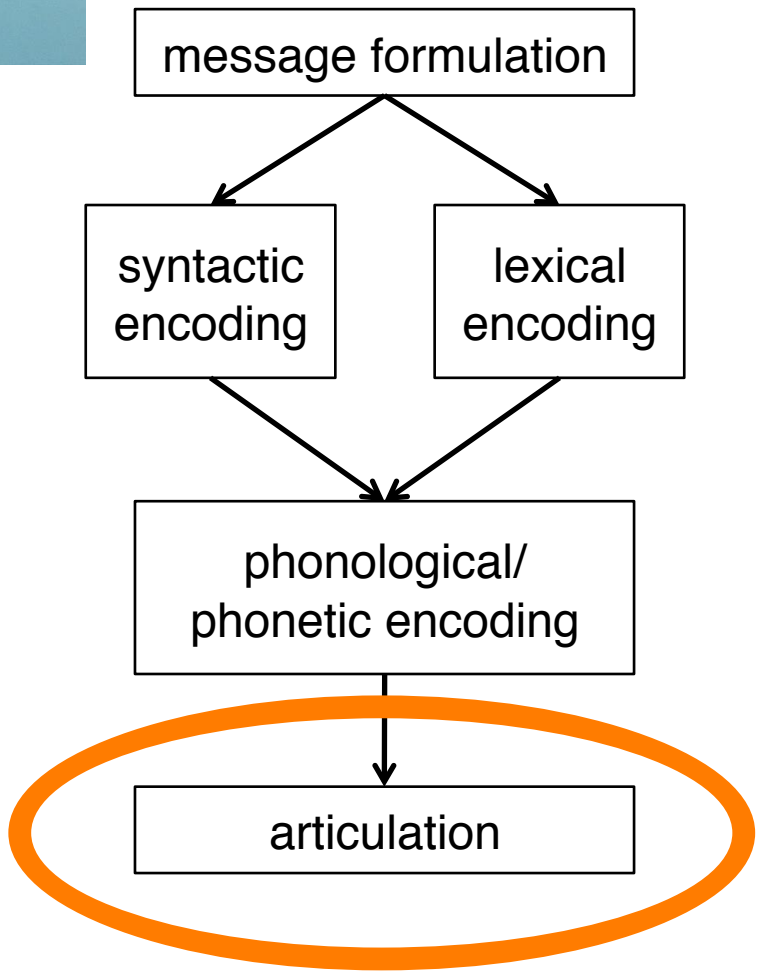
Is speech production organized for robust and efficient information transfer?

Esteban Buz

BCS lunch talk — 3/26/2014

patient: HUCKLEBERRY HOUND
agent: DOG
action: CHASE



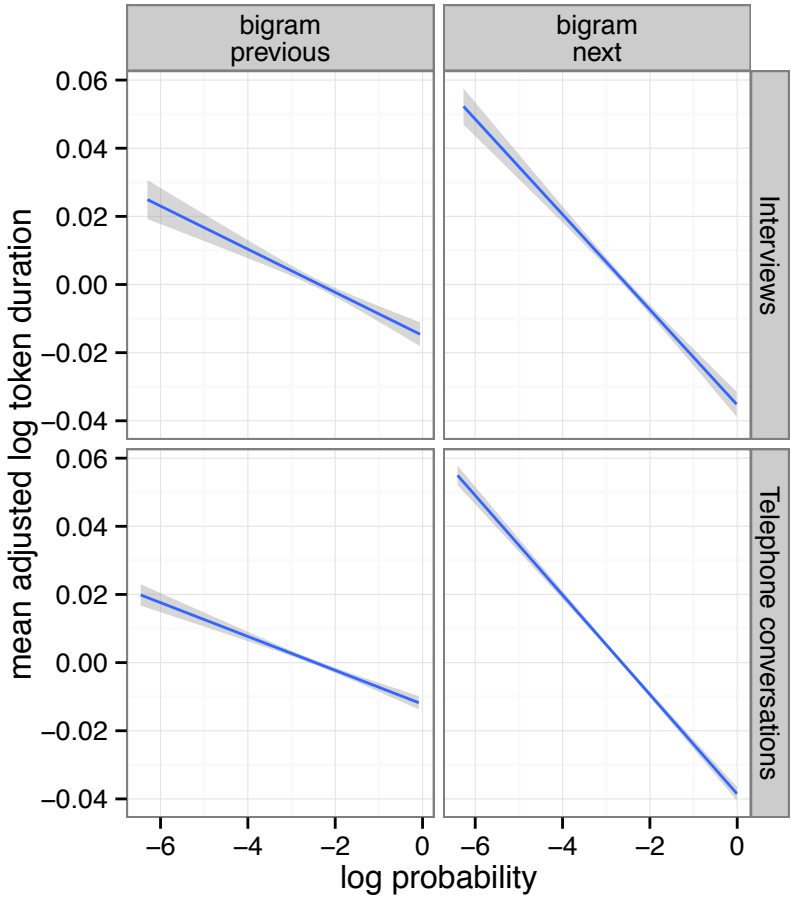
active or passive?
dog chase hound
hound chased by dog

Lexical choice?
dog
or
bulldog
or
English bulldog

stress/syllabification?
/ˈhʌk əl bɛ ri/
or
/hək əl bɛ ri ˈhaʊnd/

speech rate?
hyper-articulation?

Articulation, according to most psycholinguists



Jaeger, Seyfarth, & Buz, in prep

- Predictable words are shorter (Aylett & Turk, 2004; Bell et al., 2009; Pluymaekers et al., 2005)
- Predictable → easier to plan → reduced articulated form (Bard et al., 2000; Gahl et al., 2012)
- **Production ease drives articulation**

Articulation: goal-directed (my view)

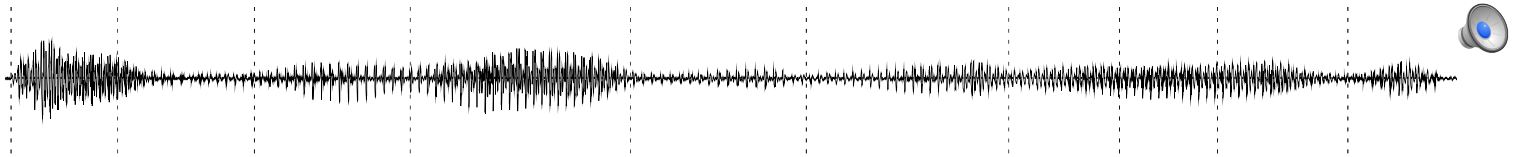
- Production planning *is* important
- **But**, being understood is also important, faithfully transmitting a message is the goal

Articulation: Parallels with motor theory

- Parallels arguments made in the motor control literature: behavior is best understood via reducing **task-relevant** error (Todorov & Jordan 2002)
- If articulation is similar to other motor systems we can view it as including a forward model plus **learning** from past task-relevant error (cf. perspective discussed in Jaeger & Ferreira, 2013)

Articulation is *fast*

In casual speech speakers utter ~5-7 syllables per second



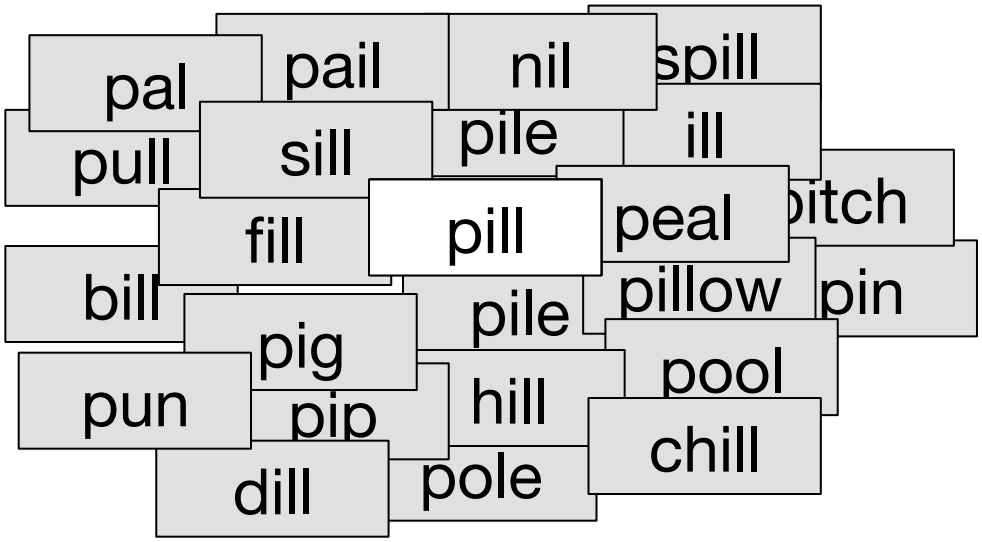
eg. 1 second of speech: 10 syllables, 21 segments, completely incomprehensible out of context.

Can we see modulations of behaviors like this?

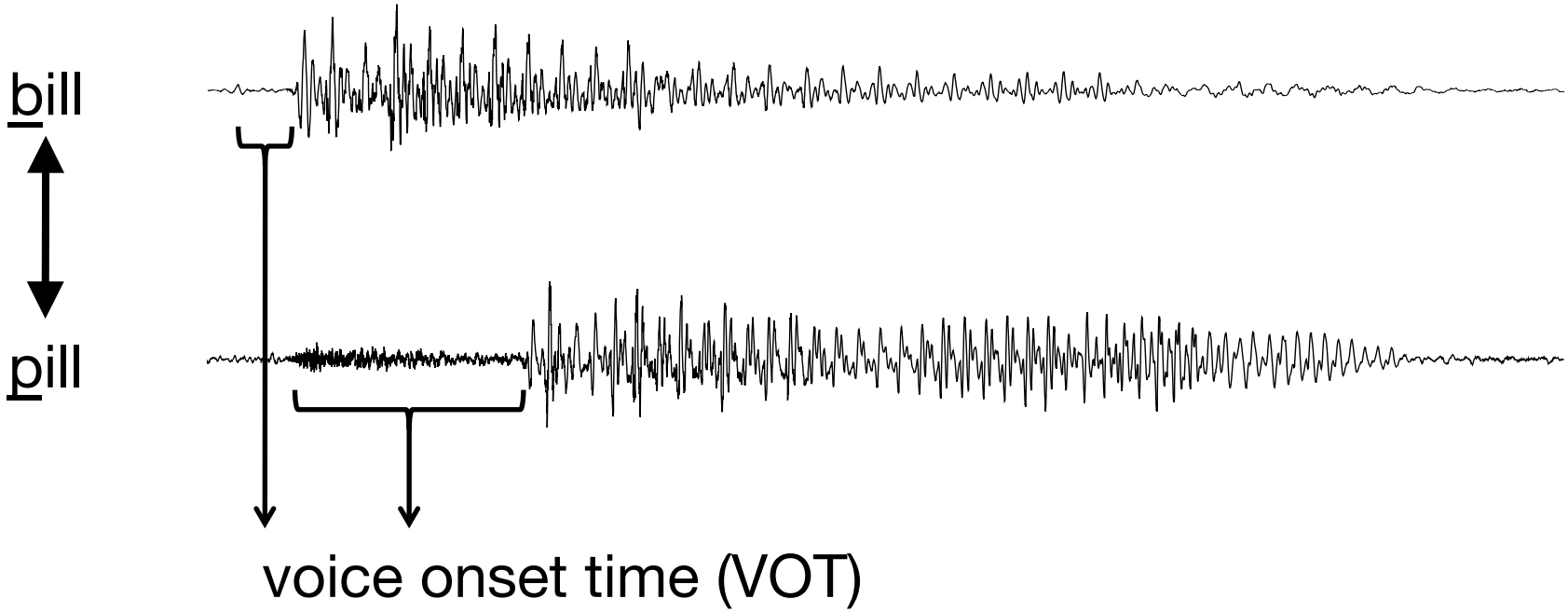
Talk overview

- Prior evidence, two views, and mine
- Three (or four) properties of articulation
- Preliminary data from a (large (partially complete)) study
- Conclusions

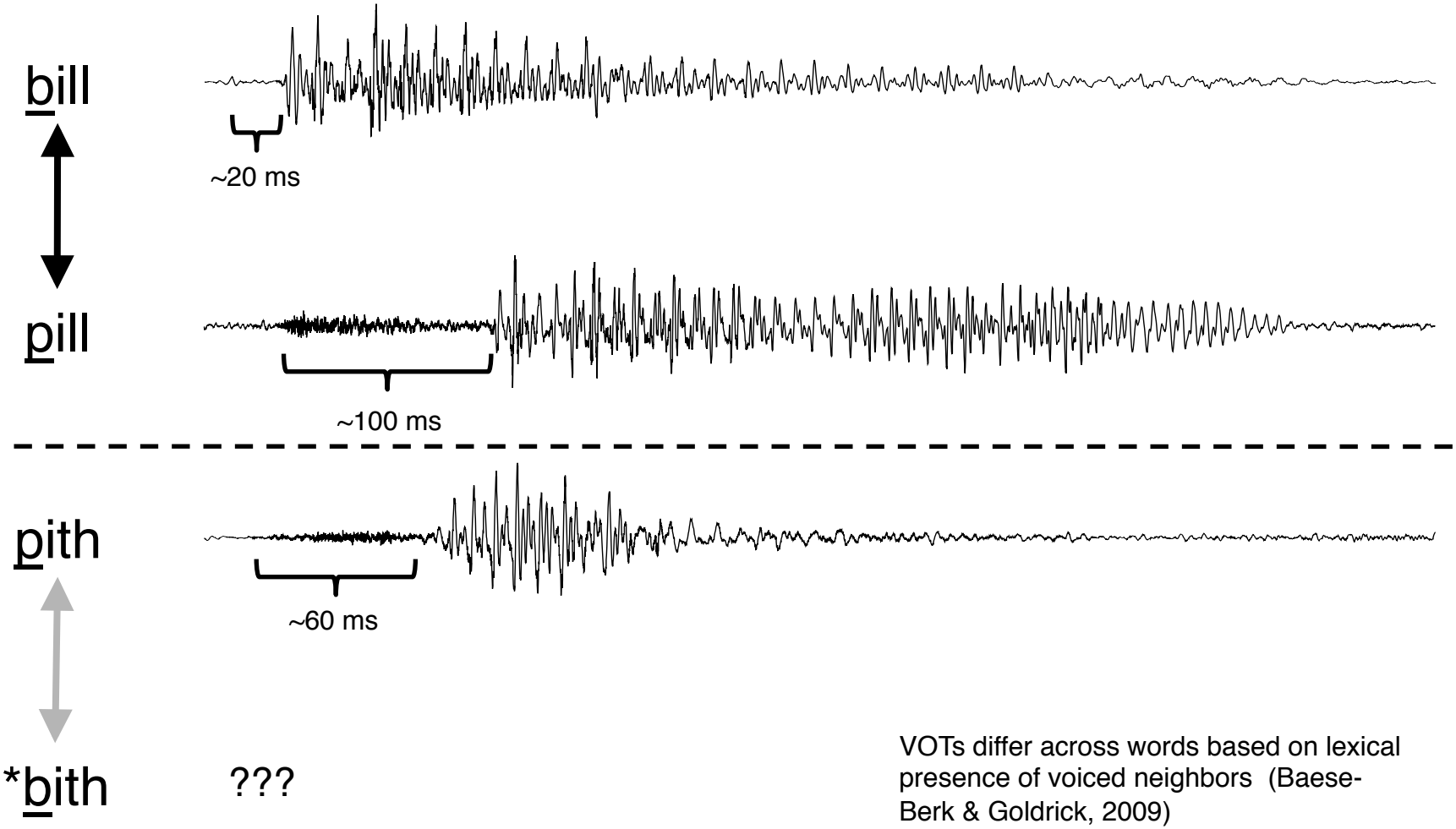
Communication and confusability



Communication and confusability



Communication and confusability



VOTs differ across words based on lexical presence of voiced neighbors (Baese-Berk & Goldrick, 2009)

Two standard accounts of these effects

- Word specific phonetics (Pierrehumbert, 2002)
  - acquired
  - not due to decisions
  - → **All lexicalized**
- Dynamics of production processes (Baese-Berk & Goldrick, 2009; Goldrick et al., 2013)
  - encoding difficulty
  - not due to communicative pressures
  - → **Production ease drives articulation**

Task-relevant modulation

How can we (I) differentiate these views?

Test case: articulation of voicing

- **Contextualized** modulations of articulation
  - Lexical account does not predict contextually driven differences

How can we (I) differentiate these views?

Test case: articulation of voicing

- **Production-ease** modulations of articulation
  - production-ease account predicts articulatory changes are solely the result of planning difficulties

How can we (I) differentiate these views?

Test case: articulation of voicing

- **Task-specific** modulations of articulation
  - to the extent that speakers learn from task-relevant error, modulations should be task-specific

How can we (I) differentiate these views?

Test case: articulation of voicing

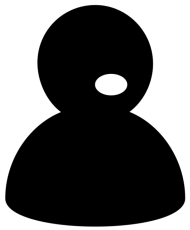
- **Learning drives** modulations of articulation
  - perceived miscommunications should result in task-relevant changes

Is the cloud's production system organized for robust communication?

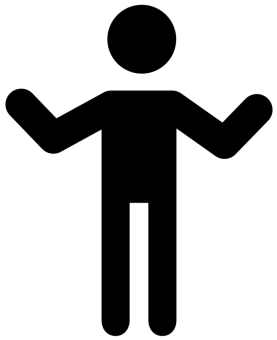
IS IT ALL LEXICAL: DOES CONTEXT MATTER?

Web-based production

bill	pill	raft

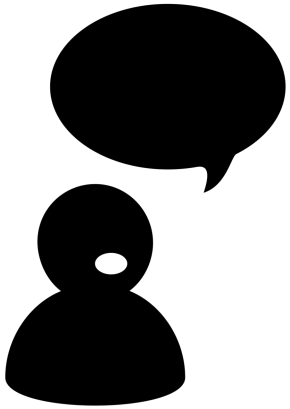


bill	pill	raft



Web-based production

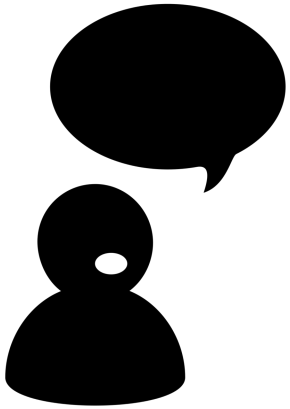
bill pill raft



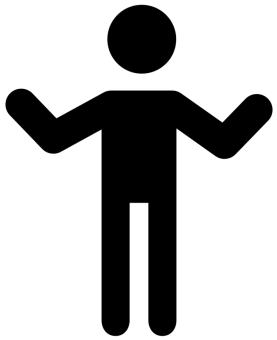
bill pill raft

Web-based production

bill pill raft



bill pill raft



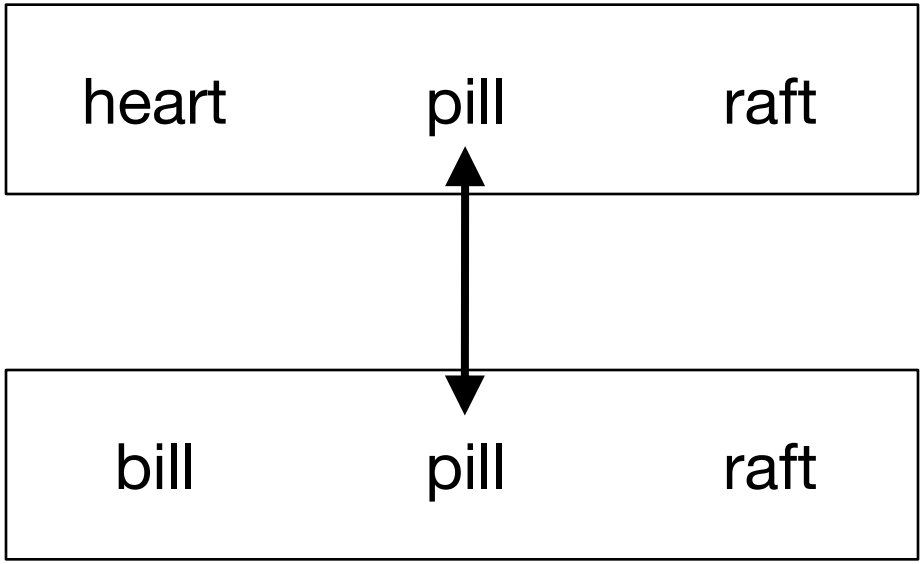
Demo

speaker task demo

Study: context and specificity

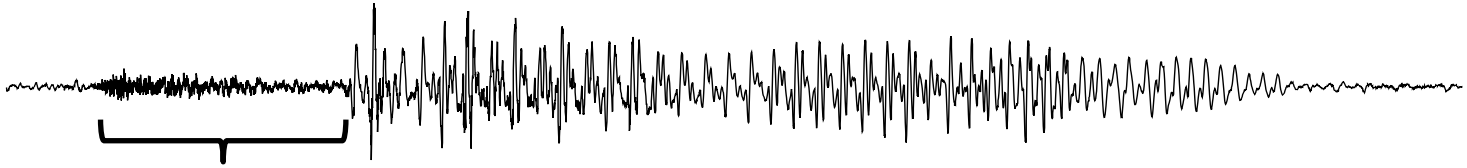
- Interactive word reading
- 36 voiceless onset critical targets
- Context manipulation: target presented with or without voiced minimal pair

- **Question: do participants change target articulation based on visual context?**

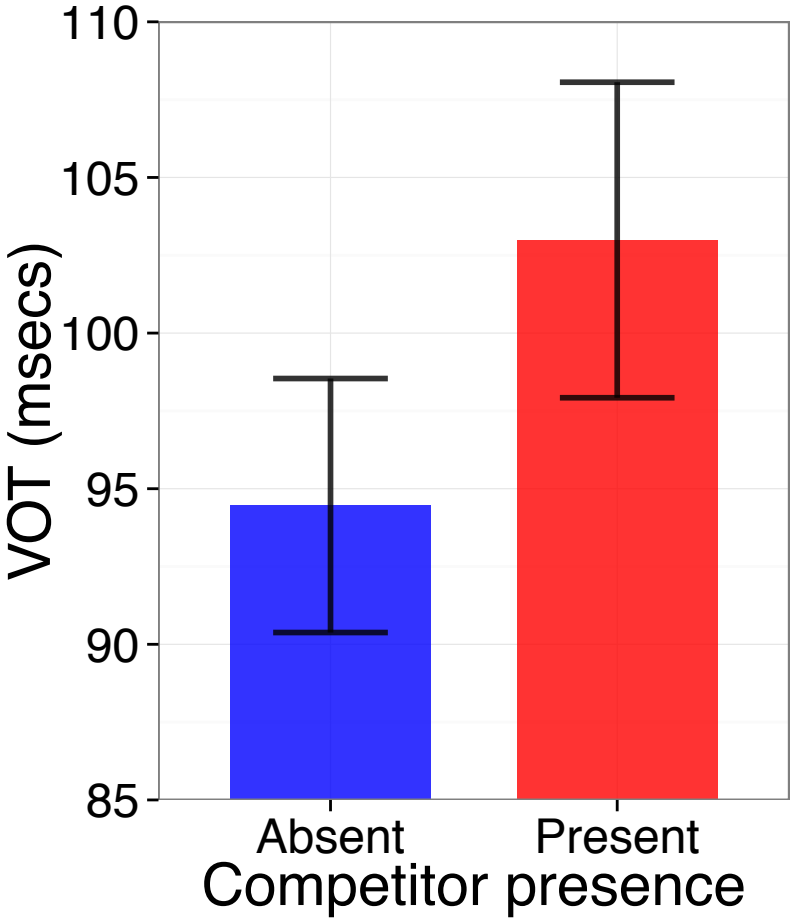


Annotations

pill



VOTs change across contexts



- VOTs are *longer* when a voiced competitor is co-present (linear mixed models, maximal RE structure, n = 10 speakers, $p < .01$)
- **Contextual confusable words produced /w more distinguishable signal (hyper-articulation)**

→ **robust communication hypothesis?**

Interim summary

- Mirrors other confusability modulated articulatory changes
  - lexically based VOT changes (Baese-Berk & Goldrick, 2009; Kirov & Wilson, 2012)
  - nhd → increases vowel distinguishability, duration and co-articulation (Munson & Solomon, 2004; Munson, 2007; Scarborough 2010; 2012
- Contextually driven affects rule out lexical account

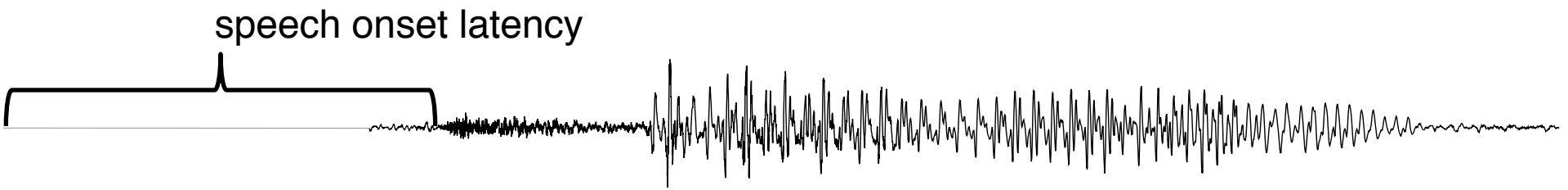
IS IT PRODUCTION EASE?

But could this after all be due to production planning?

- Modulations of articulation are only driven by production ease (Baese-Berk & Goldrick, 2009; Bell et al., 2009; Gahl et al., 2012)

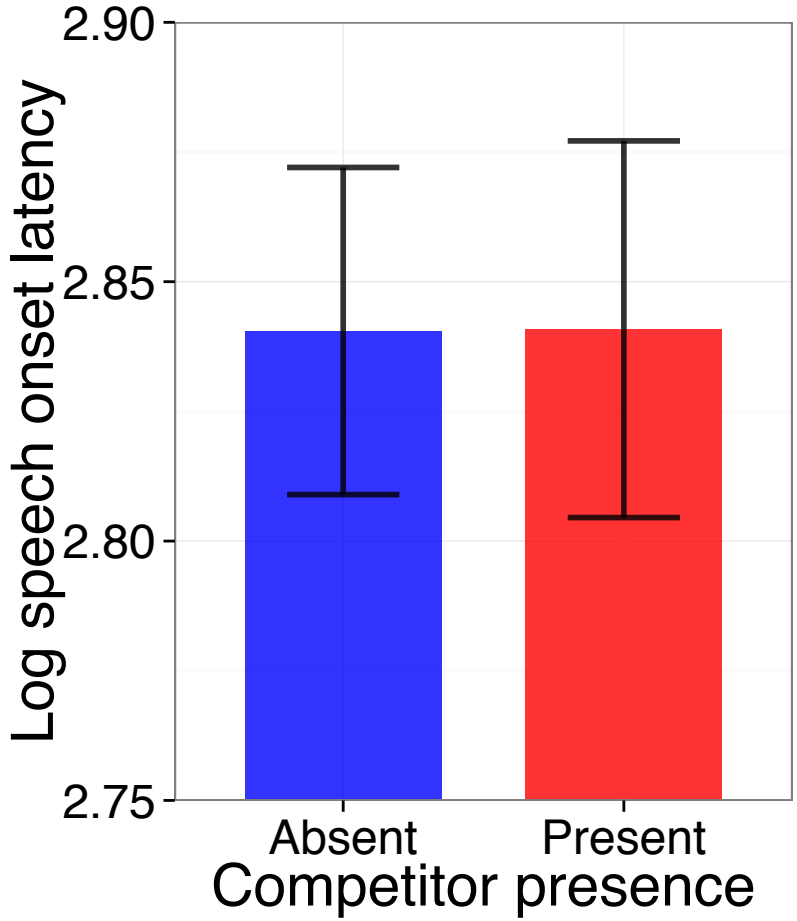
- Differences in VOTs are the result of differences in planning difficulty (Baese-Berk & Goldrick, 2009)

Speech onset latencies as a measure of planning difficulty



- **Predictions:** a) Speech onset latencies should be longer for more confusable targets and b) latencies should predict VOTs

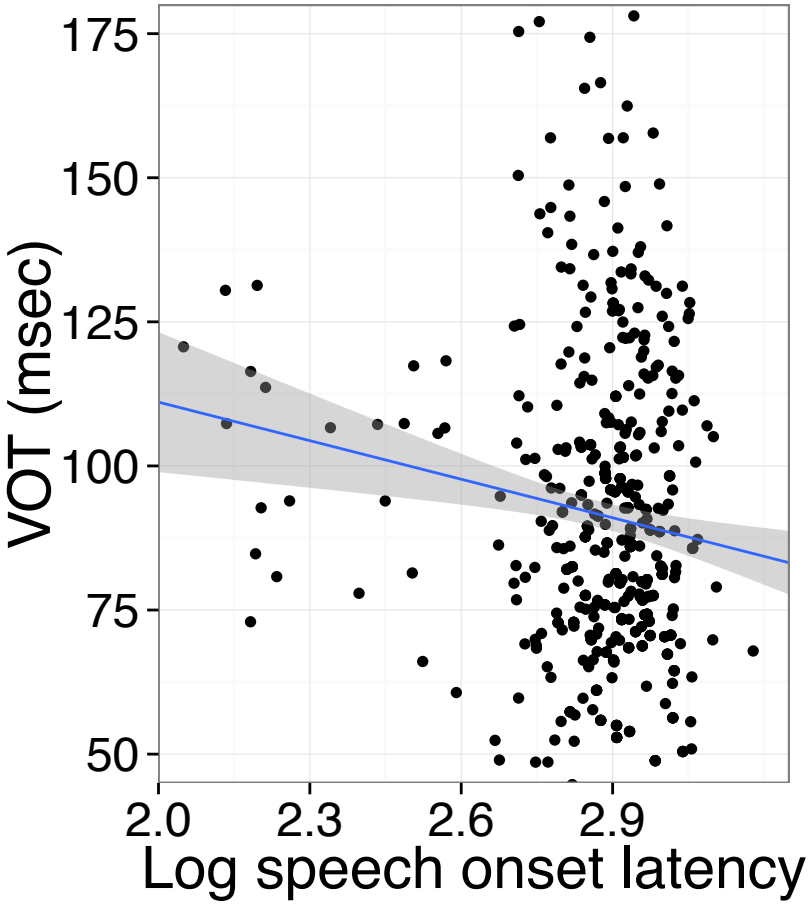
Production planning does not explain hyper-articulation



- Speech onset latencies **are equivalent** with or without competitor co-present ($p \approx .9$)

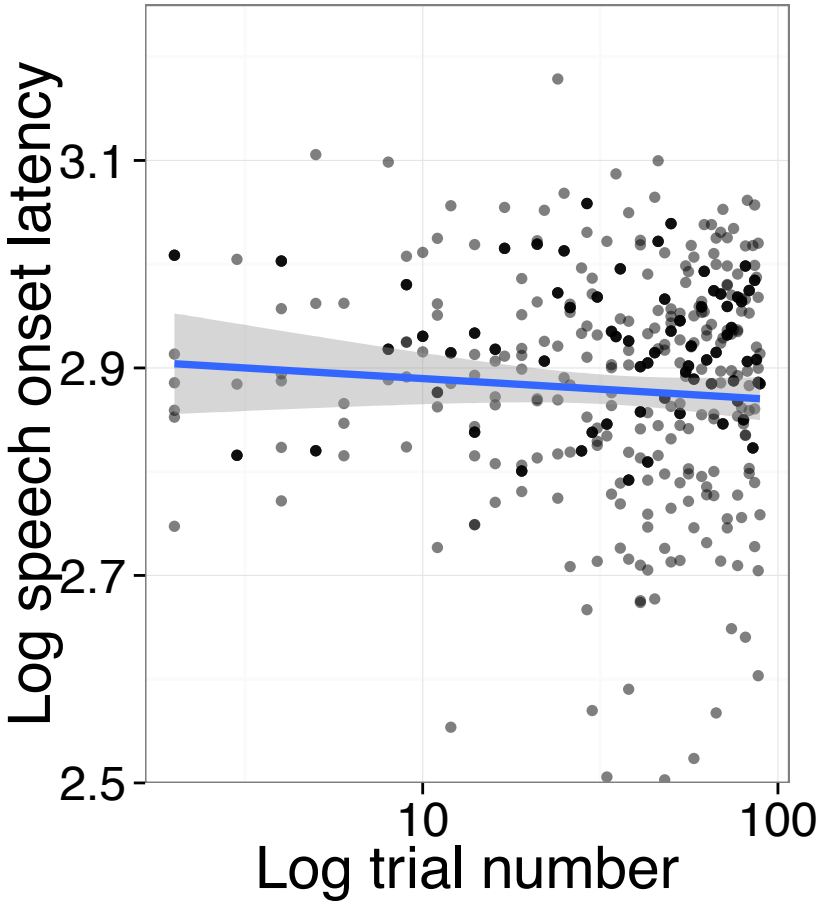
→ **NOT** production-ease account

Production planning *does not explain hyper-articulation*



- If anything, we find longer latencies predict *shorter* VOT

Latencies don’t change over time



- Latencies are not significantly different
- Taken with VOT trial-data can’t be accounted for by production-ease accounts

Interim summary

- Ruled out lexical account: context matters
- Ruled out production-ease: context doesn't affect planning difficulty (latencies)

- Tentative support for the idea that articulatory modulations are task-relevant

- Next, …
- **Specificity:** To the extent that speakers can infer the task-relevant hyper-articulation that is informative?
- **Learning:** Does hyper-articulation change as a result of task-relevant error (perceived miscommunication)?

HOW SPECIFIC IS HYPER-ARTICULATION?

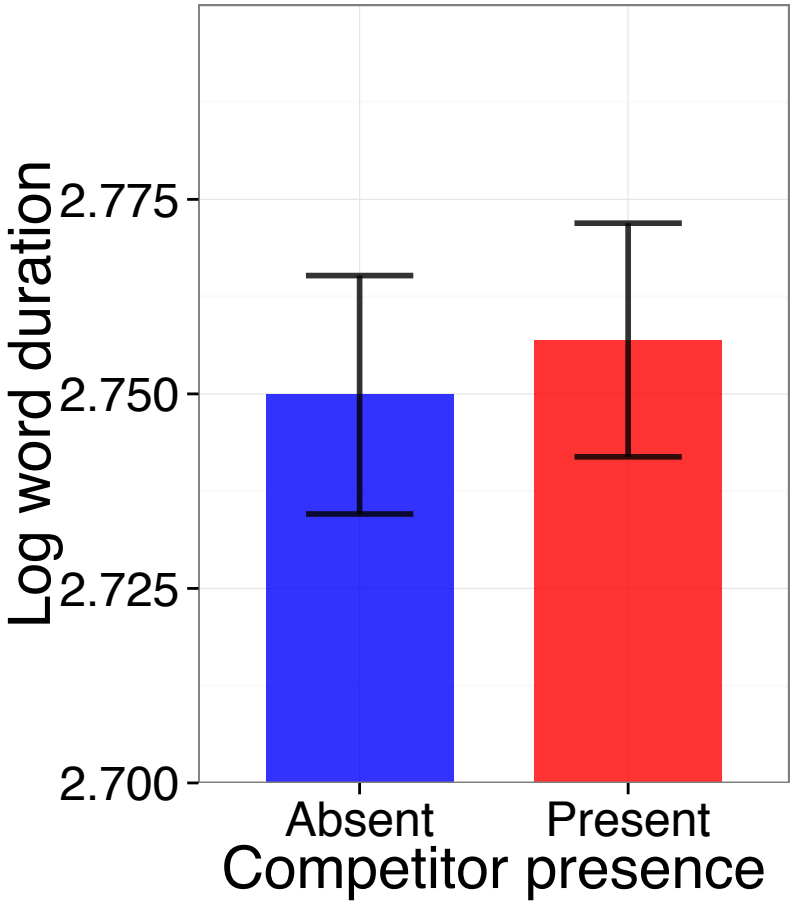
How specific is the hyper-articulation?

- Interestingly, both specific and coarse-grained measures of hyper-articulation have been found to correlate with (contextual) confusability
  - nhd → co-articulation (Scarborough, 2010, 2012)
  - nhd → vowel dispersion (Gahl et al 2012; Munson & Solomon, 2004; Munson 2007, Scarborough, 2010, 2012)
  - nhd → duration (Buz & Jaeger, 2012b, 2013; Gahl et al 2012; Munson & Solomon, 2004; Munson 2007)
  - predictability → duration (Buz & Jaeger, 2012a, 2012b)
- **But,** specific and coarse-grained measures tend to be correlated

So, how specific is the hyper-articulation?

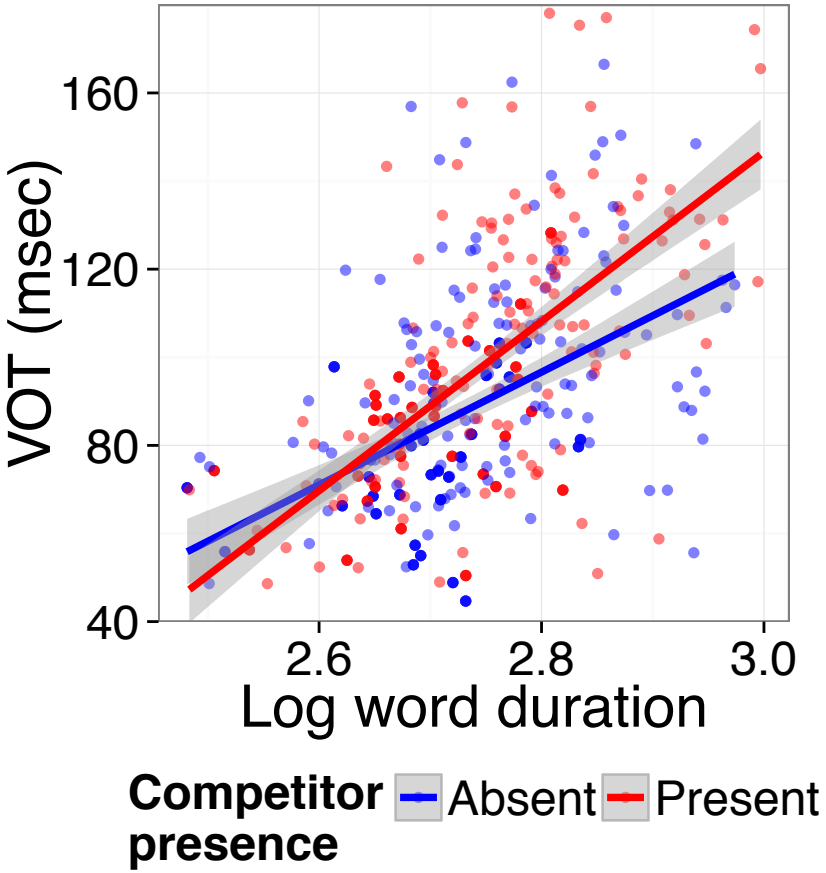
- To the extent that speakers infer (and learn from) the ***task-relevant*** error then modulations should be ***specific***: the part of the signal that carries contextually relevant contrastive information (/p/ in *pill* when *bill* is present) should be hyper-articulated

Word durations do not change across contexts



- Word durations not significantly different with or without competitor co-present
  ($p \approx .4$)

Hyper-articulation is specific to segment that carries distinguishing information



- Word durations do predict VOTs ($p < .01$)
- But, **context effect holds beyond duration** ($p < .01$)
- Interaction: VOTs are increasingly longer for longer word durations ($p < .01$)

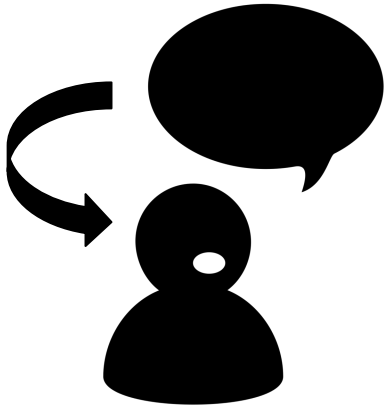
Interim summary

- Hyper-articulation is specific to the contextually relevant distinguishing segment
  - VOTs preferentially lengthened
  - Overall word length doesn't significantly differ across context

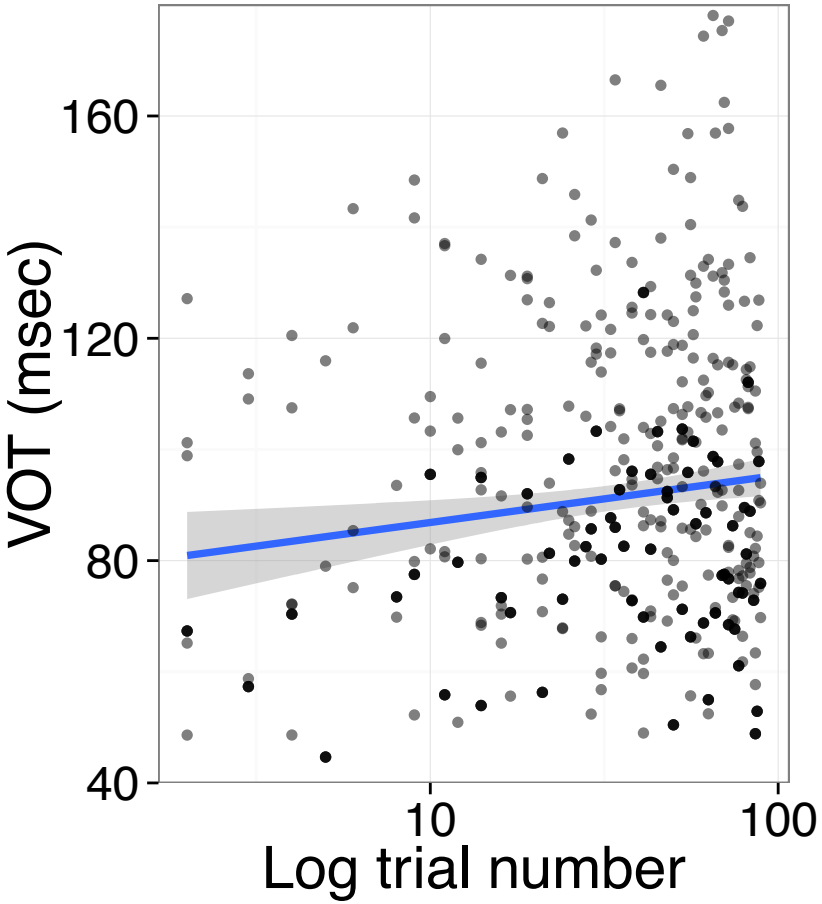
DO WE *LEARN* WHEN AND HOW TO HYPER-ARTICULATE?

Feedback (speaker internal vs external)

- Speakers in previous study ‘evaluate’ their own speech



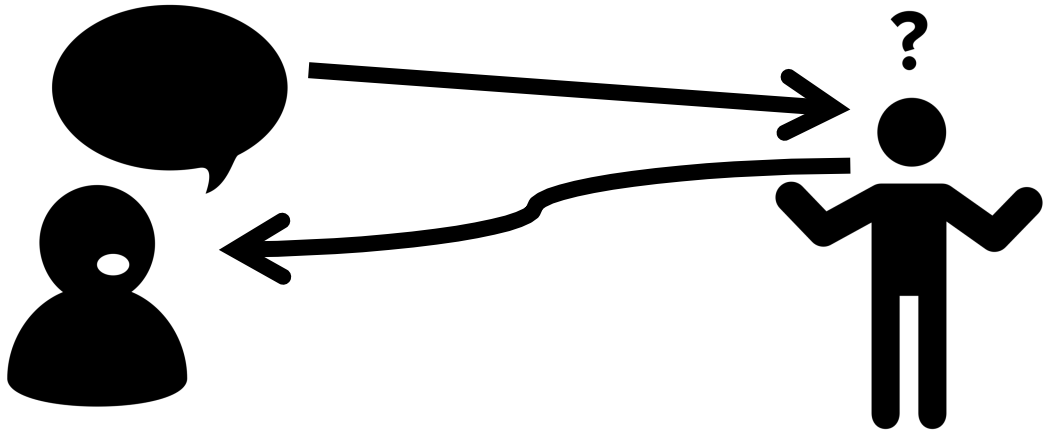
Case for learning?



- VOTs are longer at the end of the experiment ($p < .01$)

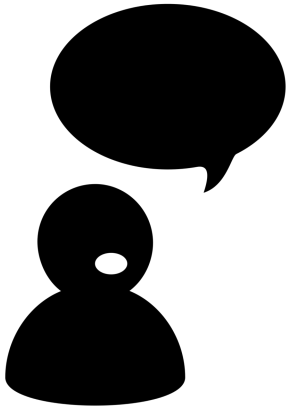
Feedback (speaker internal vs external)

- But the goal is transmit information to someone else, not yourself



Web-based production w/ feedback

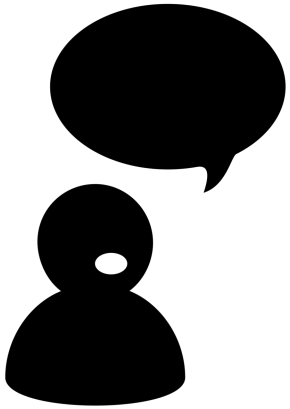
bill pill raft



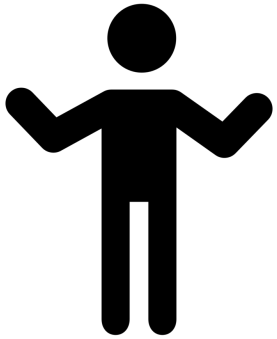
bill pill raft

Web-based production w/ feedback

bill pill raft

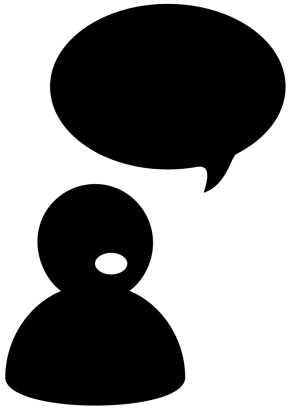


bill pill raft

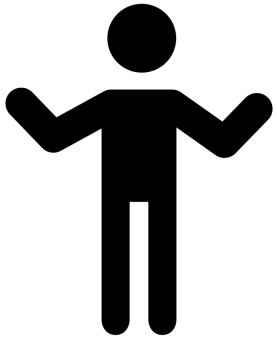


Web-based production w/ feedback

bill pill raft

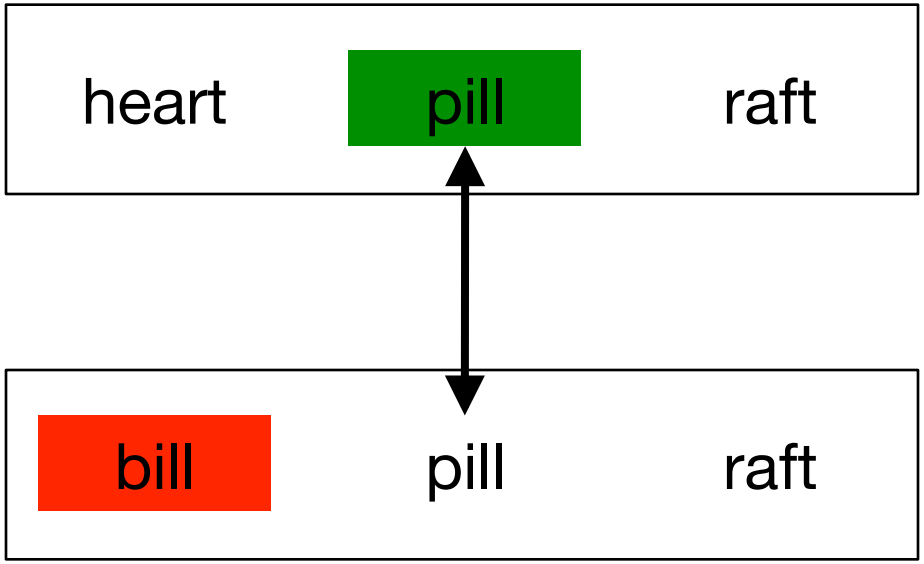


bill pill raft

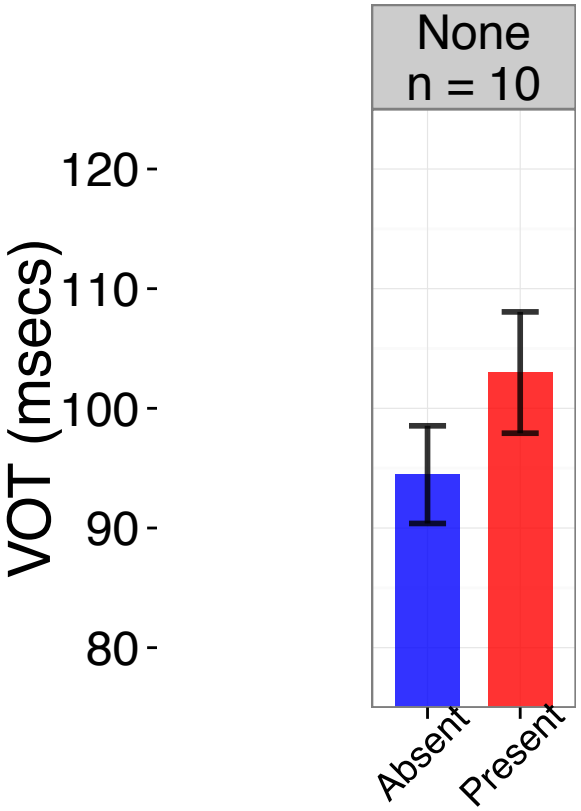


Study: Manipulating feedback

- Same stimuli as the prior study
  - Interactive word reading
  - 36 voiceless onset critical targets
  - Context manipulation: target presented with or without voiced minimal pair
- Participants split into two feedback groups: Positive and Mixed
  - Positive: partner was always right
  - Mixed: partner was occasionally wrong (5 critical trials; 1 filler)
- **Question: do participants change target articulation based on feedback?**



Preliminary VOT results

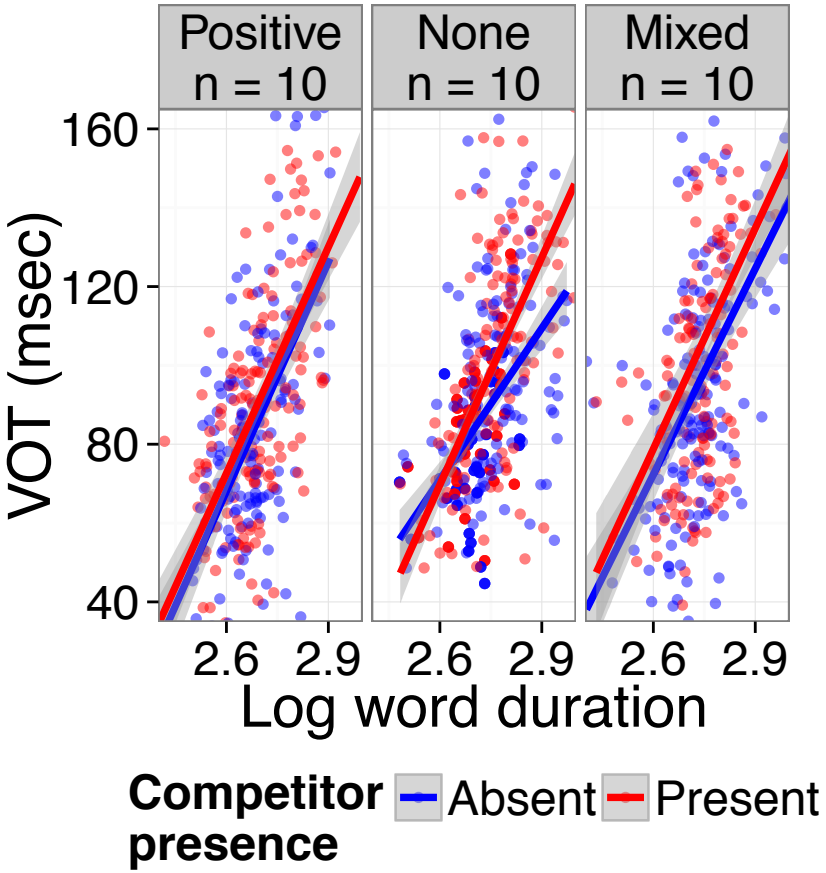


- Feedback nominally affects VOT, though not significantly ($p \approx .2$)
- Slight interaction, speakers with mixed feedback show greater difference across context ($p \approx .2$)

→ Participants adjust articulations based on **perceived** communicative success?

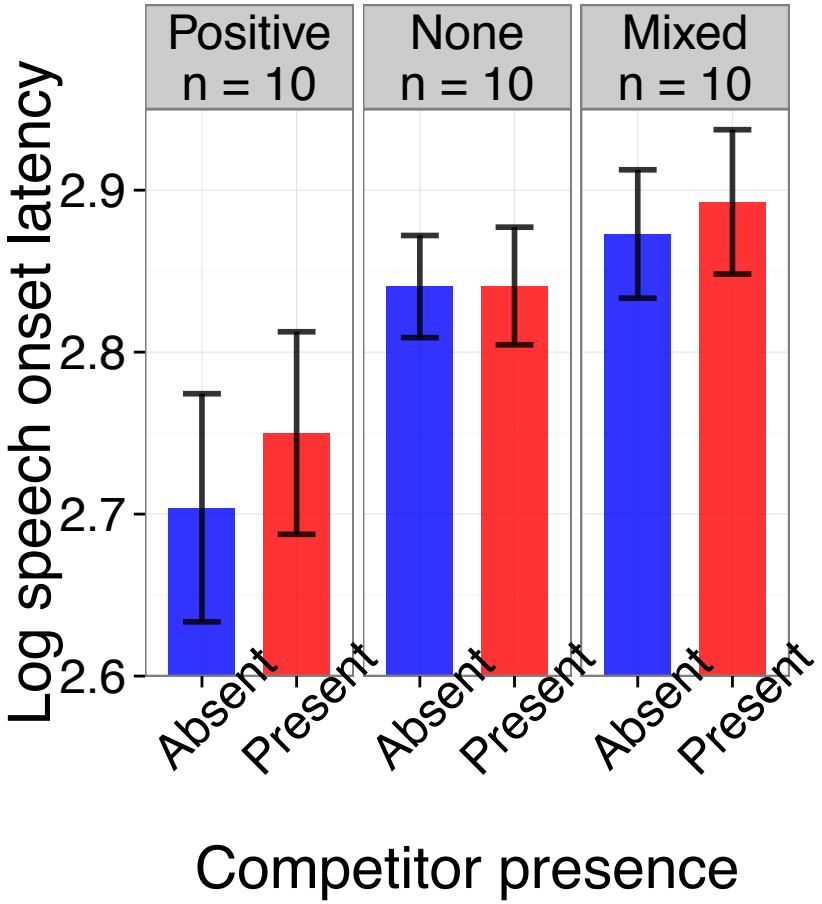
(**not predicted** by production-ease account)

Hyper-articulation is specific to segment that carries distinguishing information



- Word durations predict VOTs ($p < .01$)
- But, **context effect holds beyond duration** ($p < .01$)
- plus marginally significant interaction: VOTs are increasingly longer for longer word durations

Preliminary latency/production difficulty results



- Speech onset latencies **are equivalent** with or without competitor co-present
- Again, no evidence that context causes production difficulty
- Argues production-ease account of VOT differences

Feedback study summary

- Replication of context effect across feedback groups
  - Participants hyper-articulate VOTs of voiceless target words in the presence of voiced competitors
- Suggestive evidence for the role of feedback
  - Participants across feedback groups had nominal differences in *overall* VOTs and possible interaction
- all tentative (!!)

Conclusions

- Context specific articulations
  - rejects other accounts

- argues for task-relevant modulations of articulation to achieve communicative goals

Potential account for predictability → reduction

- Speakers provide the input for the next generation of learners
- Production biases toward efficient communication may affect what is learned (see also communicative biases in acquisition, Fedzechkina et al., 2012; 2013)

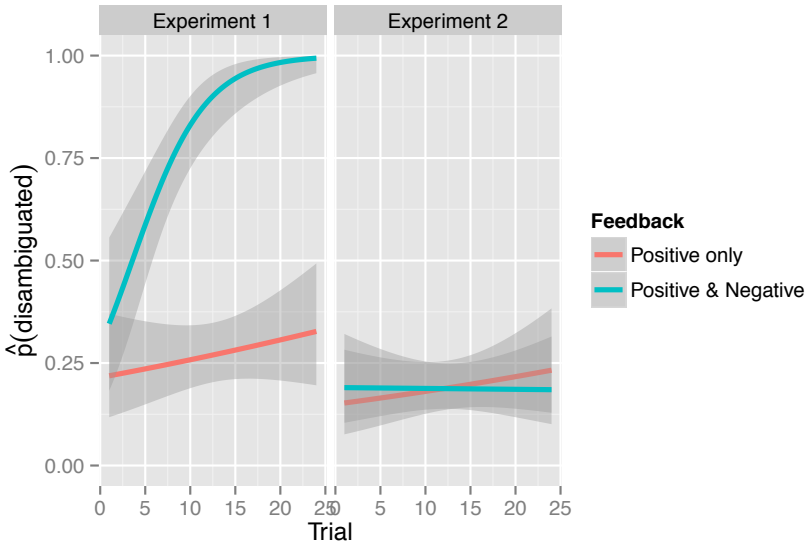
- Possible route toward generating communicatively efficient properties of language structure
  - Lexical evidence for efficient communication (Piantadosi et al., 2011)
  - Phonetic/phonological (Graff & Jaeger, 2009)
  - Language entropy (Qian & Jaeger, 2012)
  - Dependency length (Gildea & Temperley, 2010)

Open questions: feedback

- Role of feedback on learning
  - What kind?
  - How much?
  - How often?

Open questions: feedback

- Our findings are consistent with other evidence of the role of feedback on syntactic choice in production



Roche, Dale, Jaeger, & Kreuz, under revision

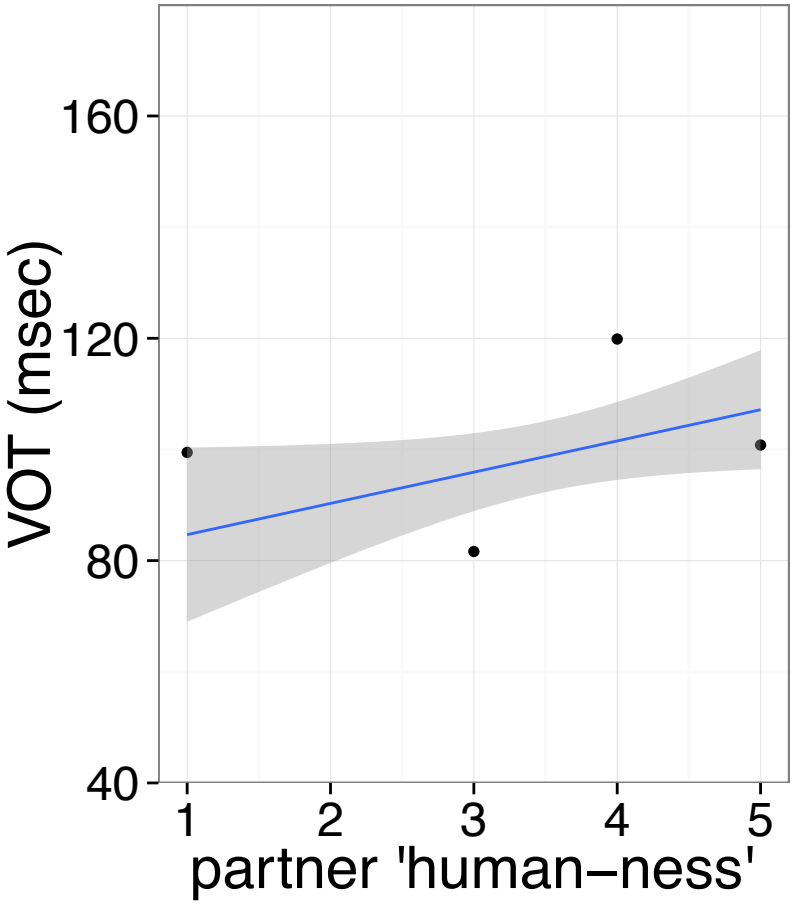
Open questions: speaker behavior

- Speaker-internal
  - What contexts result in articulatory changes?
  - Do speakers make ‘arbitrary’ articulatory changes?

Open questions: comprehension

- How do these differences affect comprehension?
  - ‘Clear’ speech does improve comprehension (Smiljanic & Bradlow, 2004)
  - Evaluating effects of feature specific modulations on comprehension like the ones presented today is as of yet un-explored
    - Two pilots run by myself have thus-far failed to find comprehension differences

Open questions: comprehension



- Do participants believe our paradigm?
  - Unprompted, majority do not say they noticed anything odd
    - Caveat: when participants are asked to rate how ‘human’ their partner is, many give relatively low ratings
    - Interestingly, this rating is mildly predictive of VOTs

Evidence for robust communication at other levels of production

- Lexical ‘choice’: lab/laboratory usage varies based on predictability (Mahowald et al., 2013)
- Reference choice supports communication (Arnold, 2008; Clark & Murphy, 1982; Tily & Piantadosi, 2009)
- Inclusion of optional words based predictability (Jaeger, 2006)
- Optional argument omission based on predictability (Kravtchenko, 2013)
- Syntactic structure results mixed: speakers do not avoid syntactic ambiguities (cf. Ferreira, 2008)

Thanks!

- Andrew Watts
- HLP RAs:
  - N. Craycraft
  - A. Grealish
  - A. Venuti
  - A. MacDonald
  - E. Rowe
- HLP/MTan/CLS family
- Funding:
  - NSF CAREER Award (NSF IIS-1150028) to TFJ
  - NIH training Grant at the University of Rochester (#T32 DC000035)