Any questions?

- Practicalities?
- Any open issues from yesterday?

Lecture 4: *talker-specific learning*

Severijnen, G. G. A., Di Donna, G., Bosker, H. R., & McQueen, J. M. (2023). Tracking talker-specific cues to lexical stress: Evidence from perceptual learning. *Journal of Experimental Psychology: Human Perception and Performance*, 49(4), 549–565. doi:10.1037/xhp0001105.

Hans Rutger Bosker

Speech Perception in Audiovisual Communication [SPEAC] lab

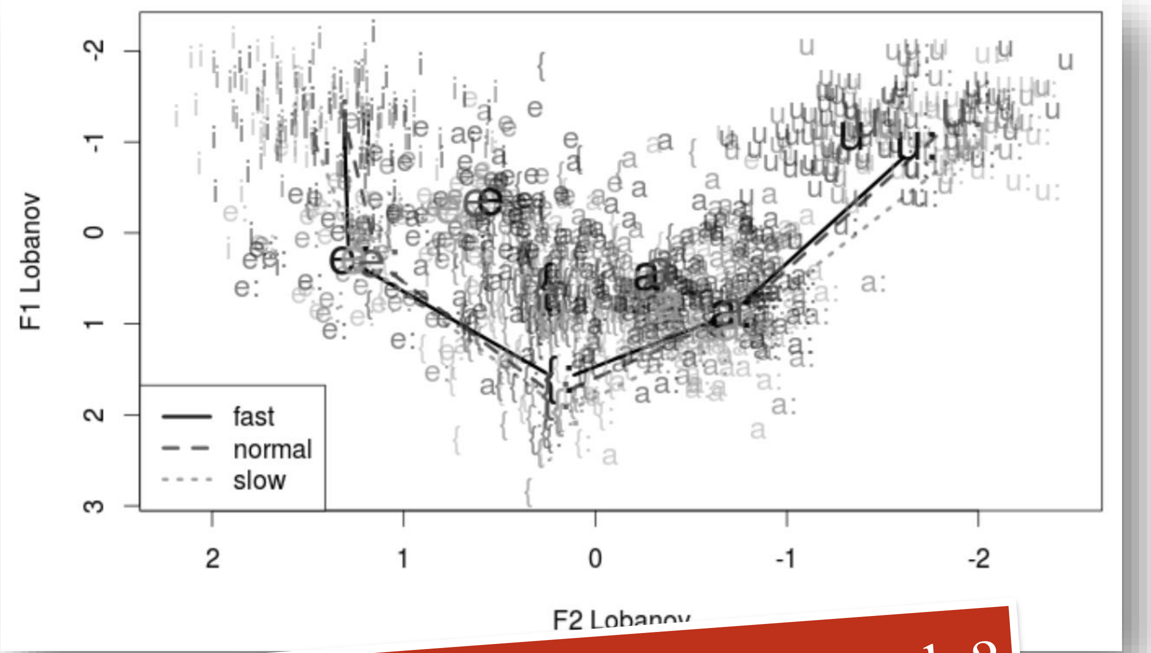
Donders Institute, Radboud University, Nijmegen, The Netherlands

https://hrbosker.github.io

hansrutger.bosker@donders.ru.nl

Variability in speech

- Remember this one?



Does the same hold for prosody?

Schulz et al. (2016)

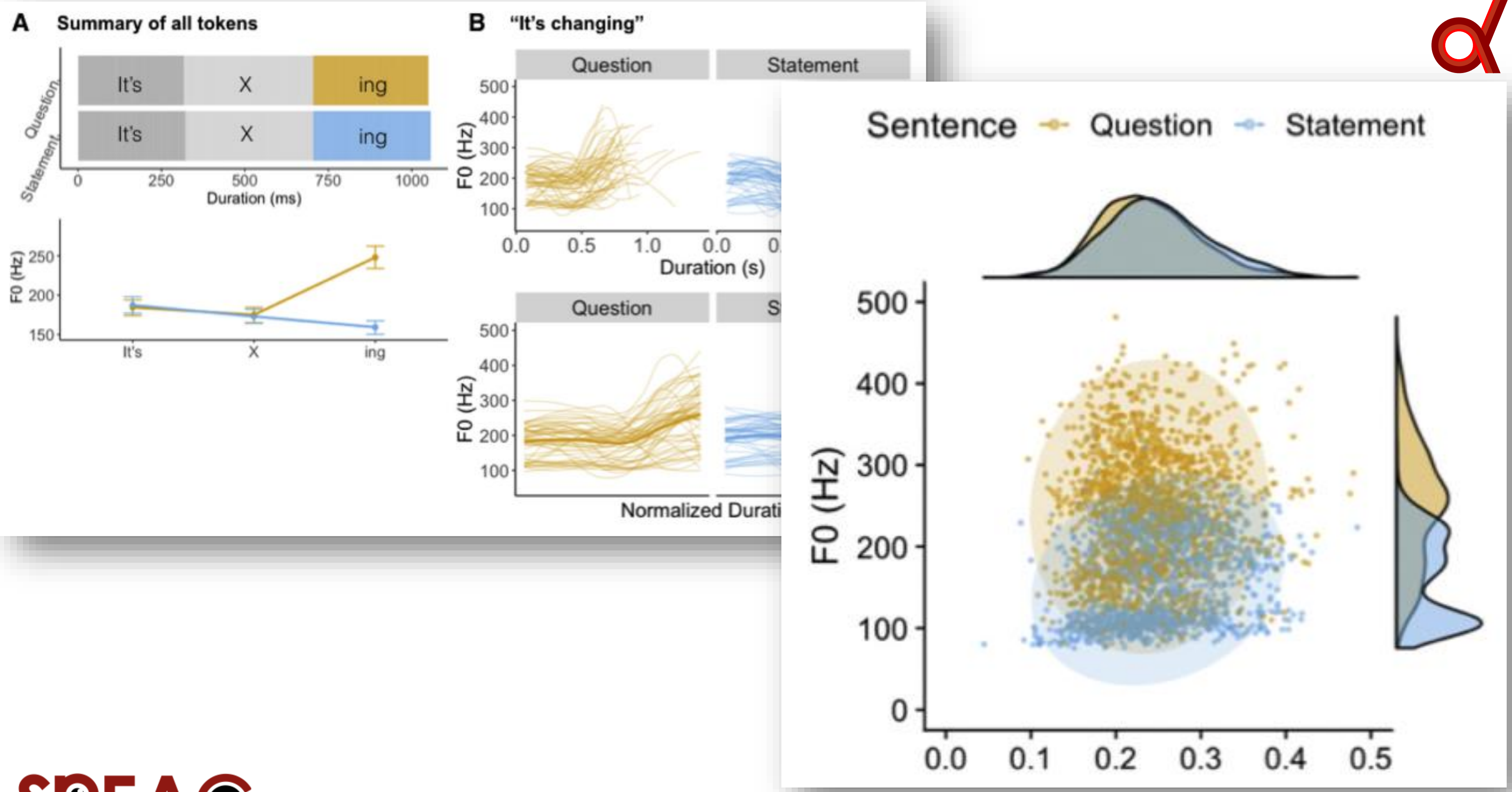
Prosody is highly variable too!

- Group-level differences?
  - Pitch height and range of male vs. female talkers
  - Talker gender and regional dialects can change your pause distributions, pitch accents, speech rate, and lexical stress.
    Clopper & Smiljanic, 2011; Arvaniti & Garding, 2007; Quené, 2008; Eriksson & Heldner, 2015
  - In Italian, women produce stressed syllables with a wider pitch range and longer syllable duration compared to men.
    Eriksson et al. (2016)
  - Your native language (e.g., a tonal language) can affect how you use *f*0 in producing lexical stress in English.
    Tseng et al. (2013)

Prosody is highly variable too!

- Individual-level differences?
  - Question vs. statement prosody

Xie et al. (2021)



Xie et al. (2021)

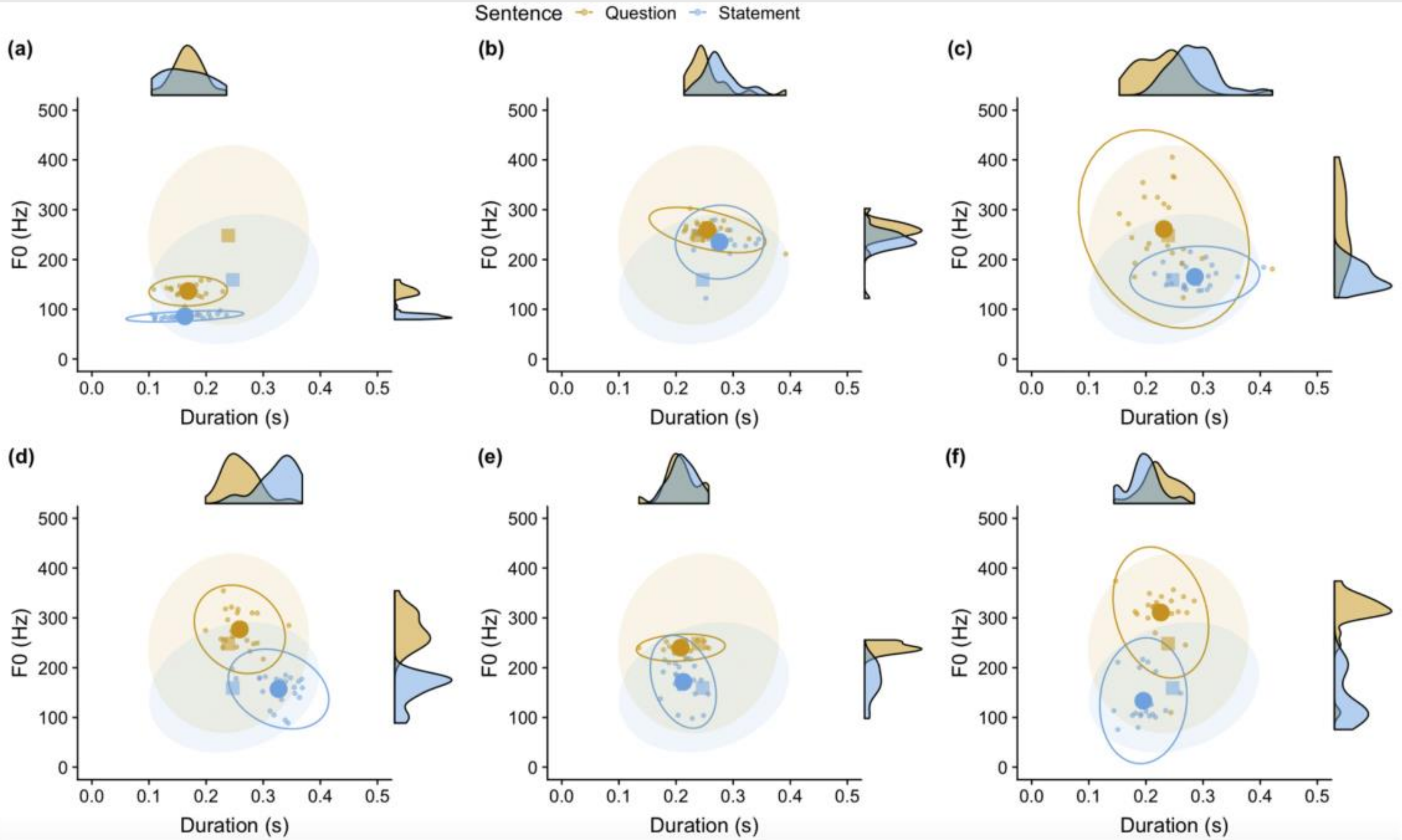


Fig. 3. Distribution of un-normalized utterance-final F0 and duration for 6 example talkers (a) –(f) from Experiment 1. Small points show individual tokens produced by the talker. Ellipses (solid lines) indicate bivariate Gaussian 95% CI of that talker's categories. Filled ellipses in the background show bivariate Gaussian 95% CI of marginal distributions of each category from Fig. 2.

Xie et al. (2021)

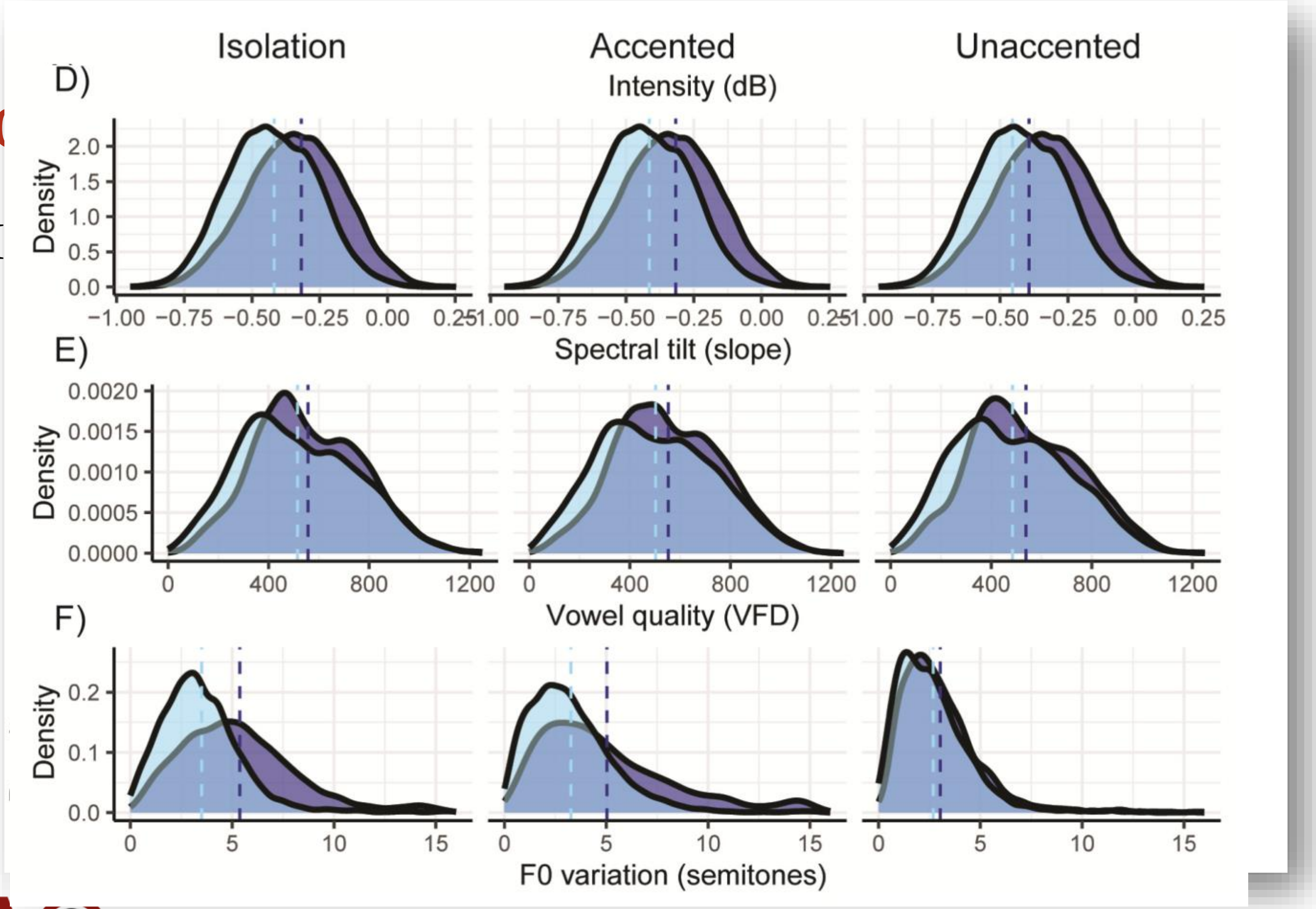
Prosody is highly variable too!

- Individual-level differences?
    - Question vs. statement prosody
    - Lexical stress
        - 40 Dutchees (20 F, 20 M) read out sentences containing ‘stress pairs’
        - e.g., “PLAto” vs. “plaTEAU”
        - Conditions: isolation, accented, unaccented

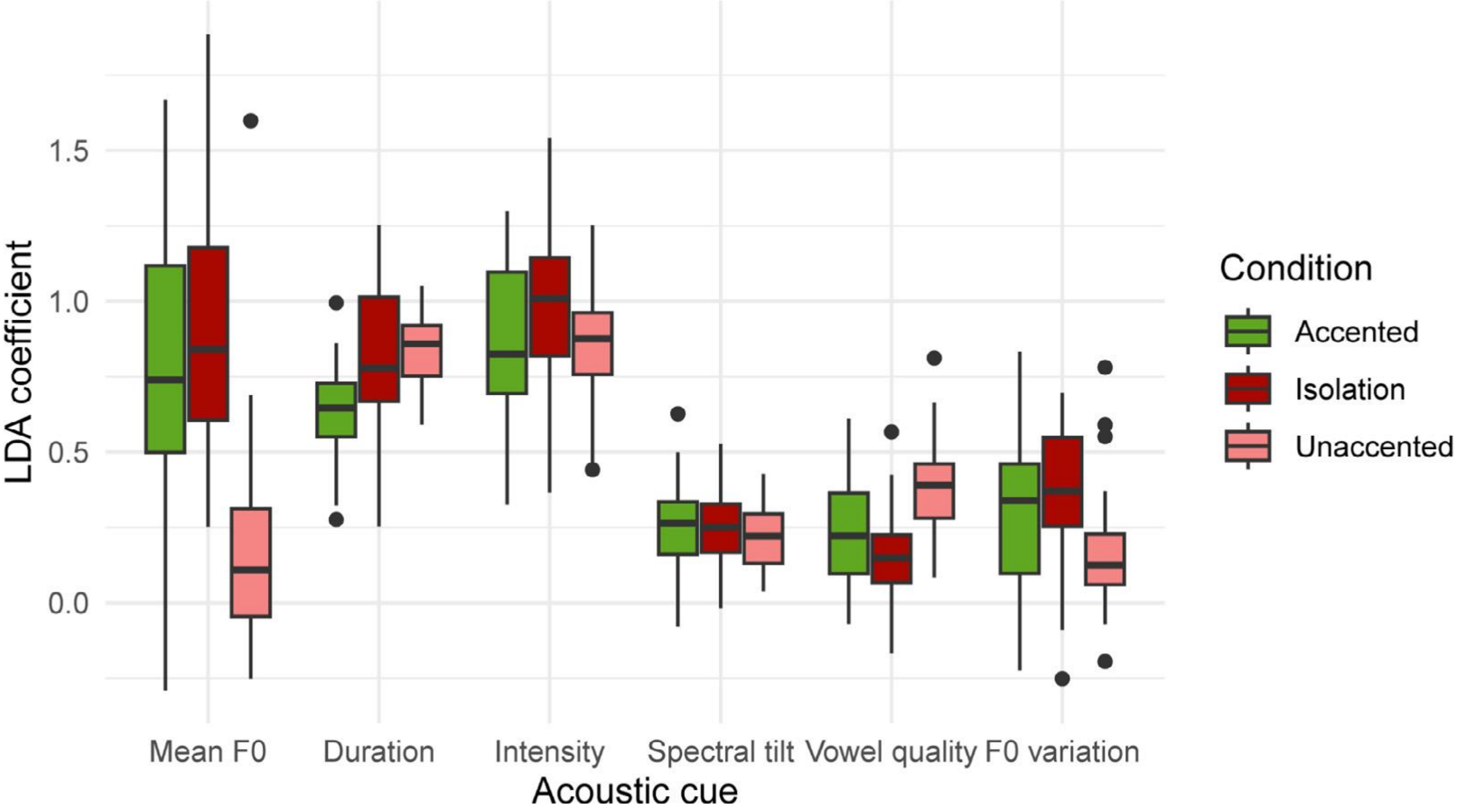
Severijnen, Bosker, & McQueen (2024, *JPhon*)

Pr

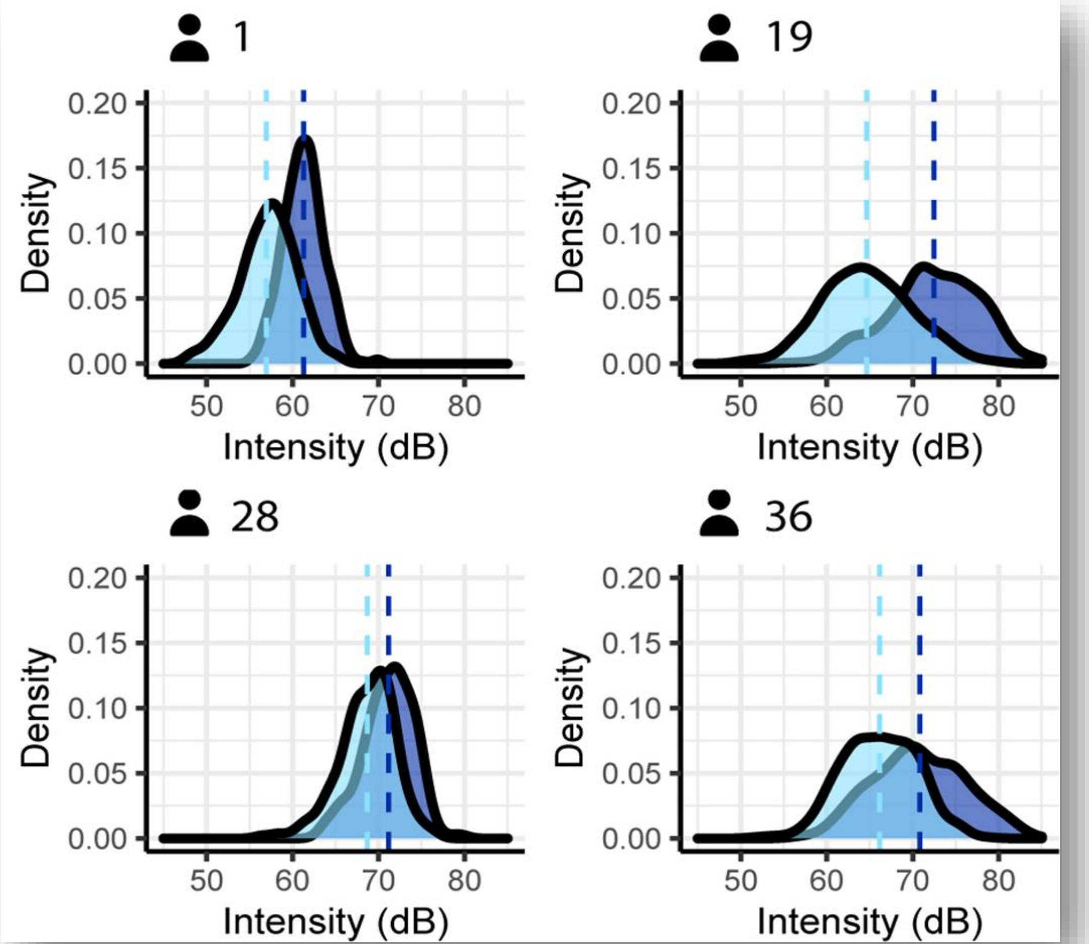
- I

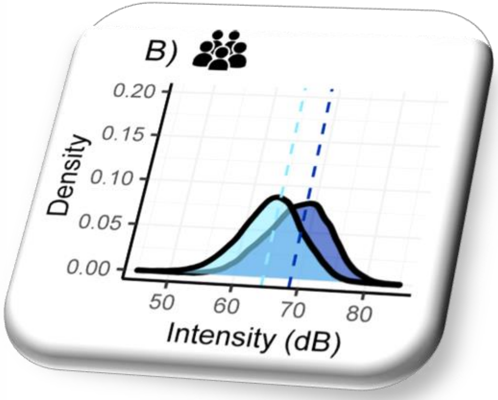


Severijnen, Bosker, & McQueen (2024, *JPhon*)

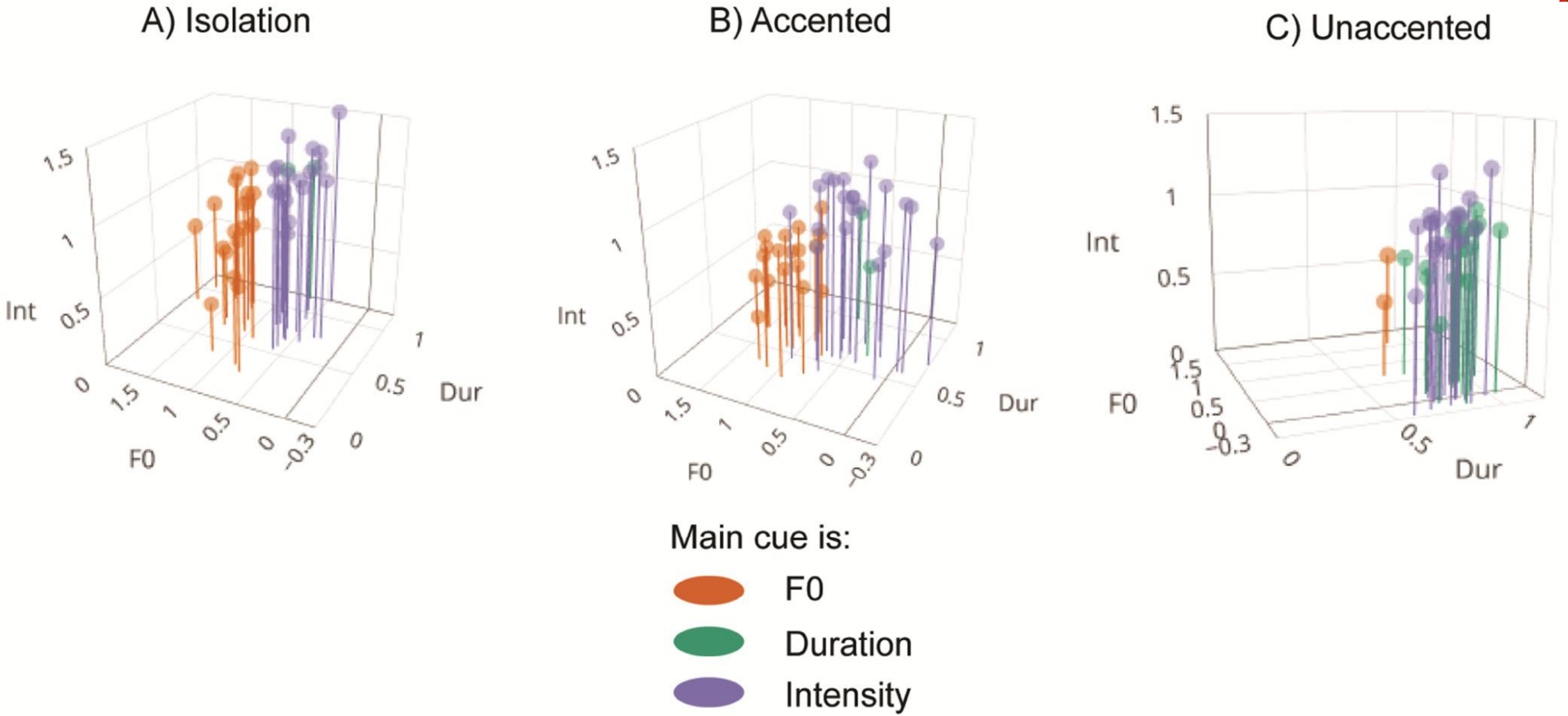


Severijnen, Bosker, & McQueen (2024, *JPhon*)





Severijnen, Bosker, & McQueen (2024, *JPhon*)



Severijnen, Bosker, & McQueen (2024, *JPhon*)

Learning about prosody

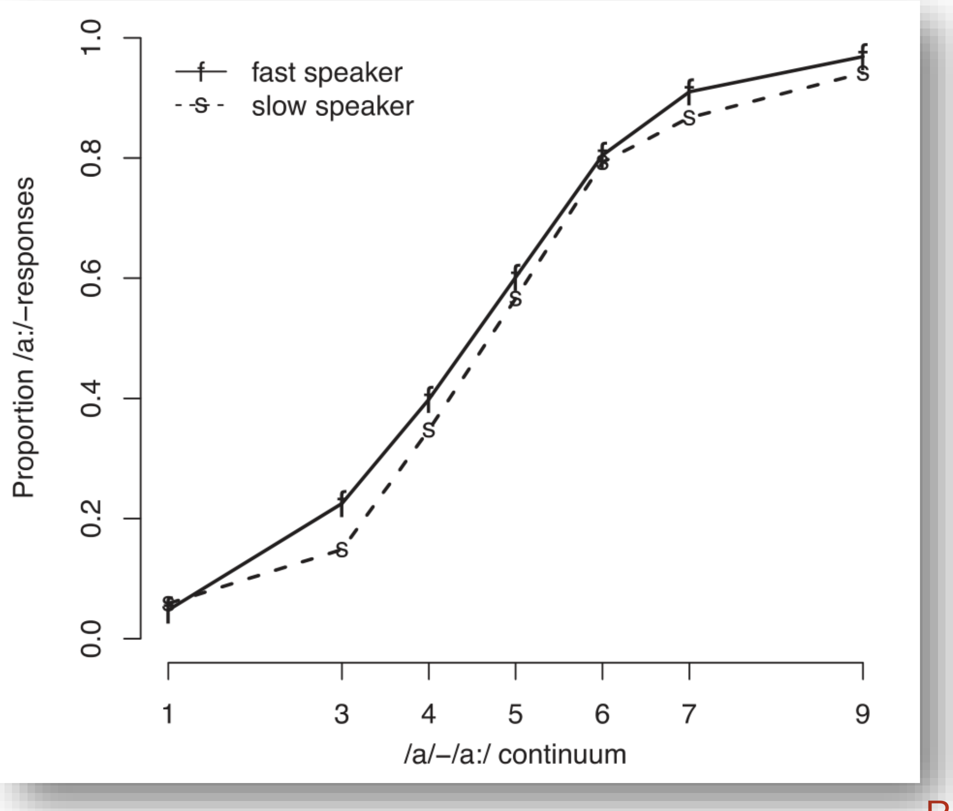
- If it’s really that bad, how do we ever manage to comprehend anything?

- Talker-specific learning to the rescue!
  - Can we demonstrate that prior knowledge about how someone speaks (i.e., talker-specific usage of prosodic cues) helps listeners comprehend new speech from that same person?
  - Typical paradigm:

exposure	>>>	**test**
learning phase		does the learning have any effect?
different for different groups		identical for both groups

Knowledge about a talker's *average* speech rate

- Rate normalization *b*
- EXPOSURE: listen t
  - Group 1: Talker
  - Group 2: Talker
- TEST: categorize /a-a

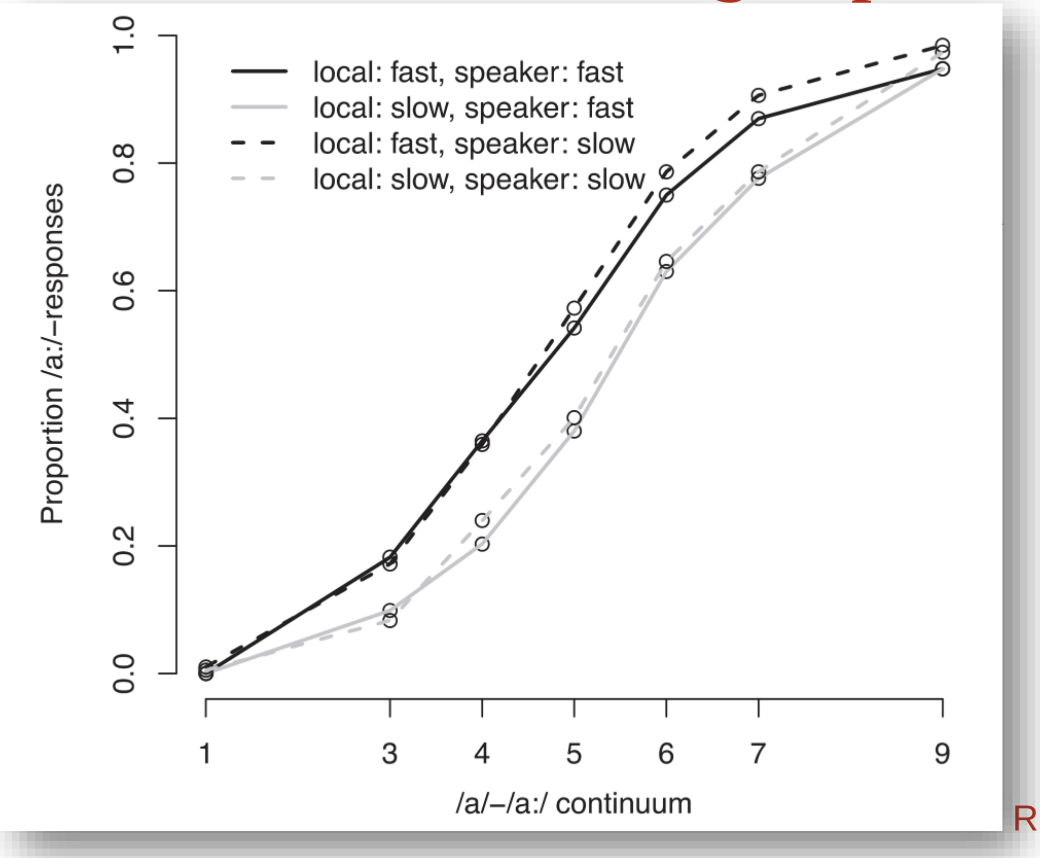


Reinisch (2016)

Knowledge about a talker’s *average* speech rate

- Rate normalizatic
- EXPOSURE: list
  - Group 1: T
  - Group 2: T
- TEST: categorize

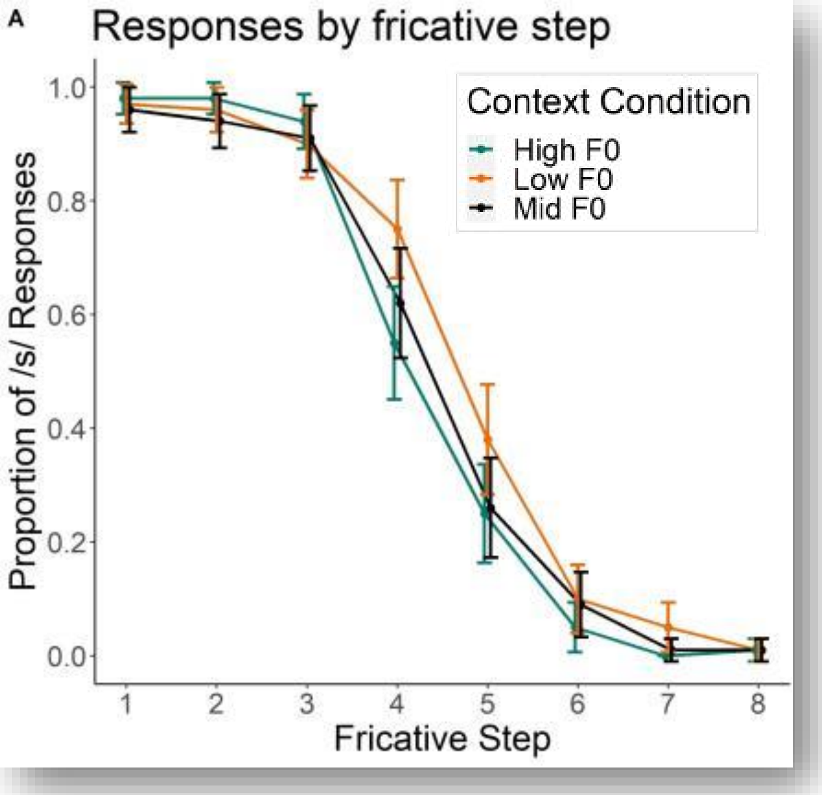
- Expt2: TEST doe
  but words in fast



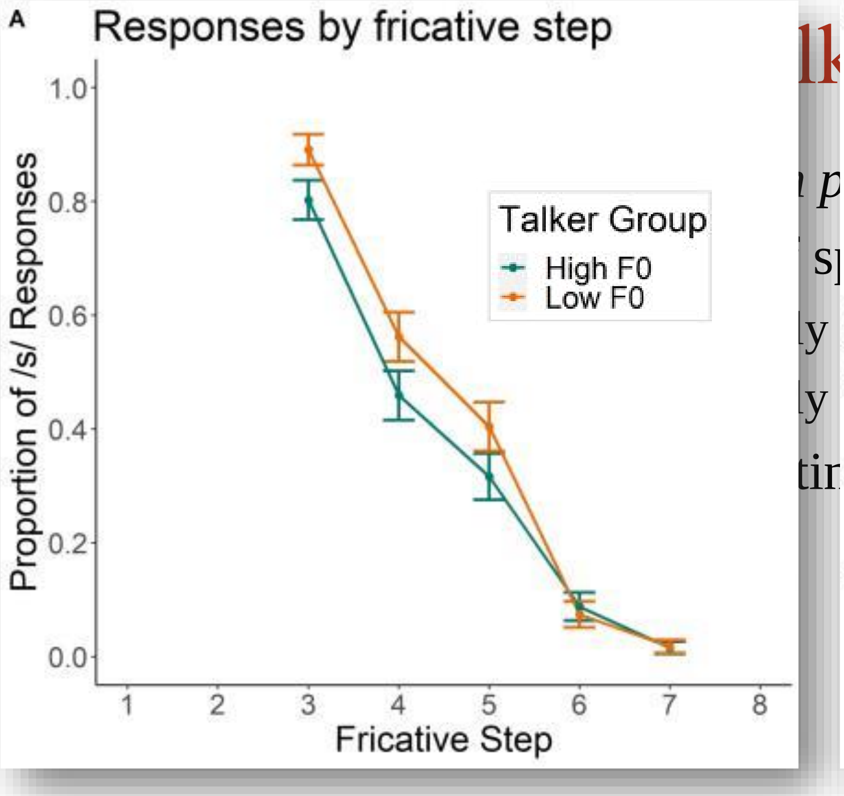
Reinisch (2016)

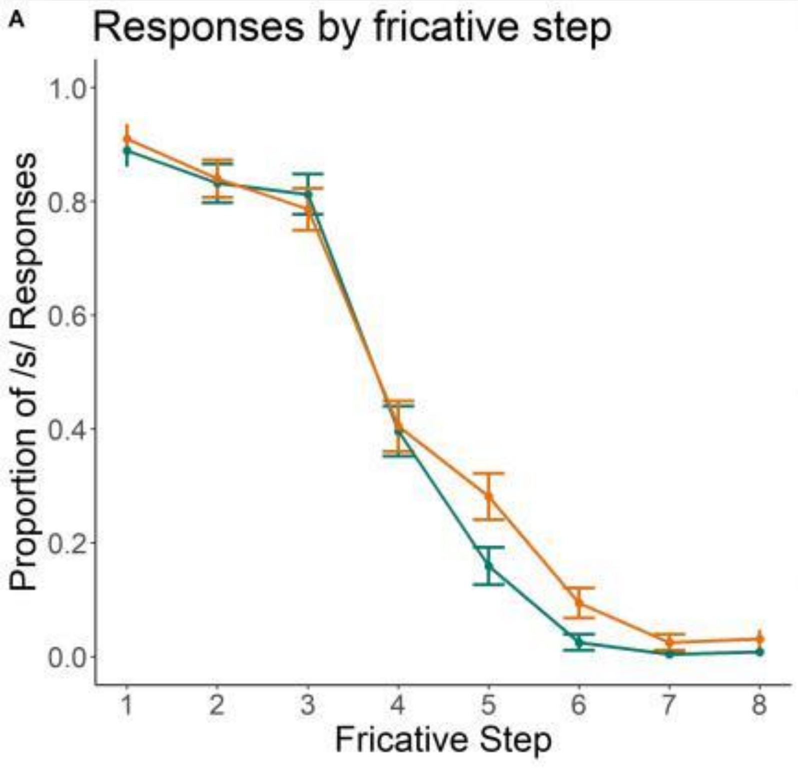
Knowledge about a talker

- Spectral normalization *based on prio*
- EXPOSURE: listen to 20min of spee
  - Group 1: Talker is habitually high
  - Group 2: Talker is habitually low
- TEST: categorize /s-ʃ/ CoG continuu



Ulusahin, Bosker, Meyer, & McQueen, *subm.*



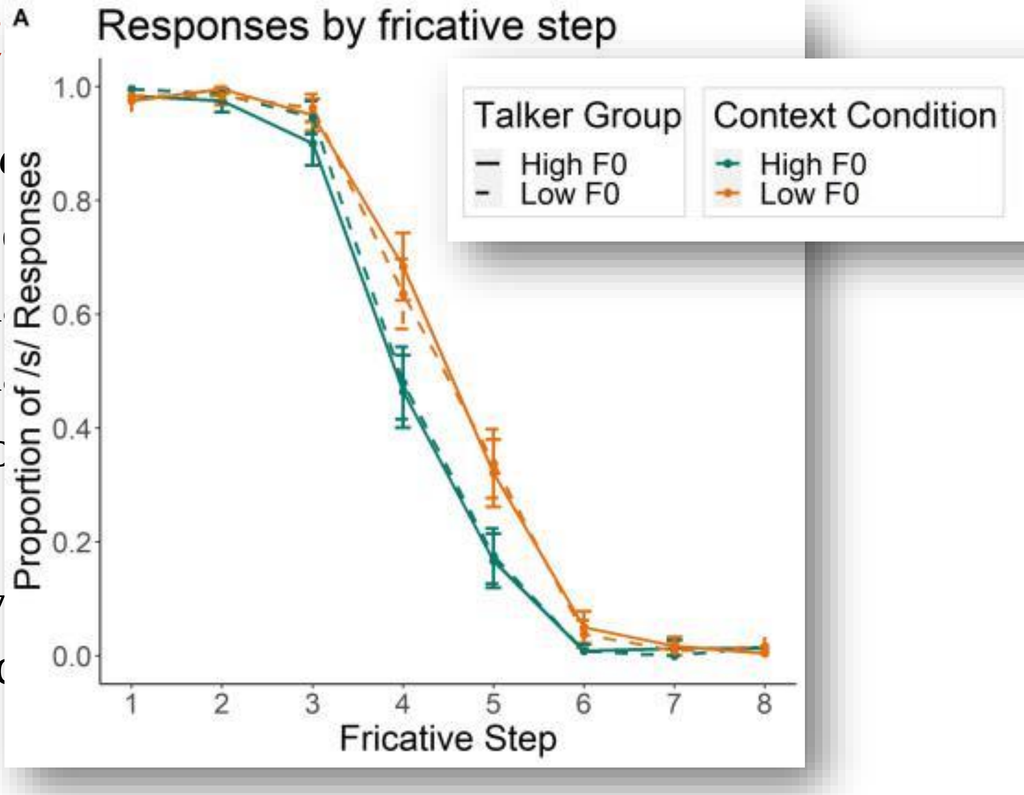


Ulusahin, Bosker, Meyer, & McQueen, *subm.*

Knowledge about a t

- Spectral normalization *based*
- EXPOSURE: listen to 20min
  - Group 1: Talker is habitu
  - Group 2: Talker is habitu
- TEST: categorize /s-ʃ/ CoG co

- Final expt: TEST does *not* inv
  but words in high vs. low-pitc



Ulusahin, Bosker, Meyer, & McQueen, *subm.*

Knowledge about a talker’s *average* prosody

- Listeners pick up on and learn about individual talker’s *average prosody*
- However, this prior knowledge is outweighed by more local information
- Crucial role for *reliability* of prior knowledge

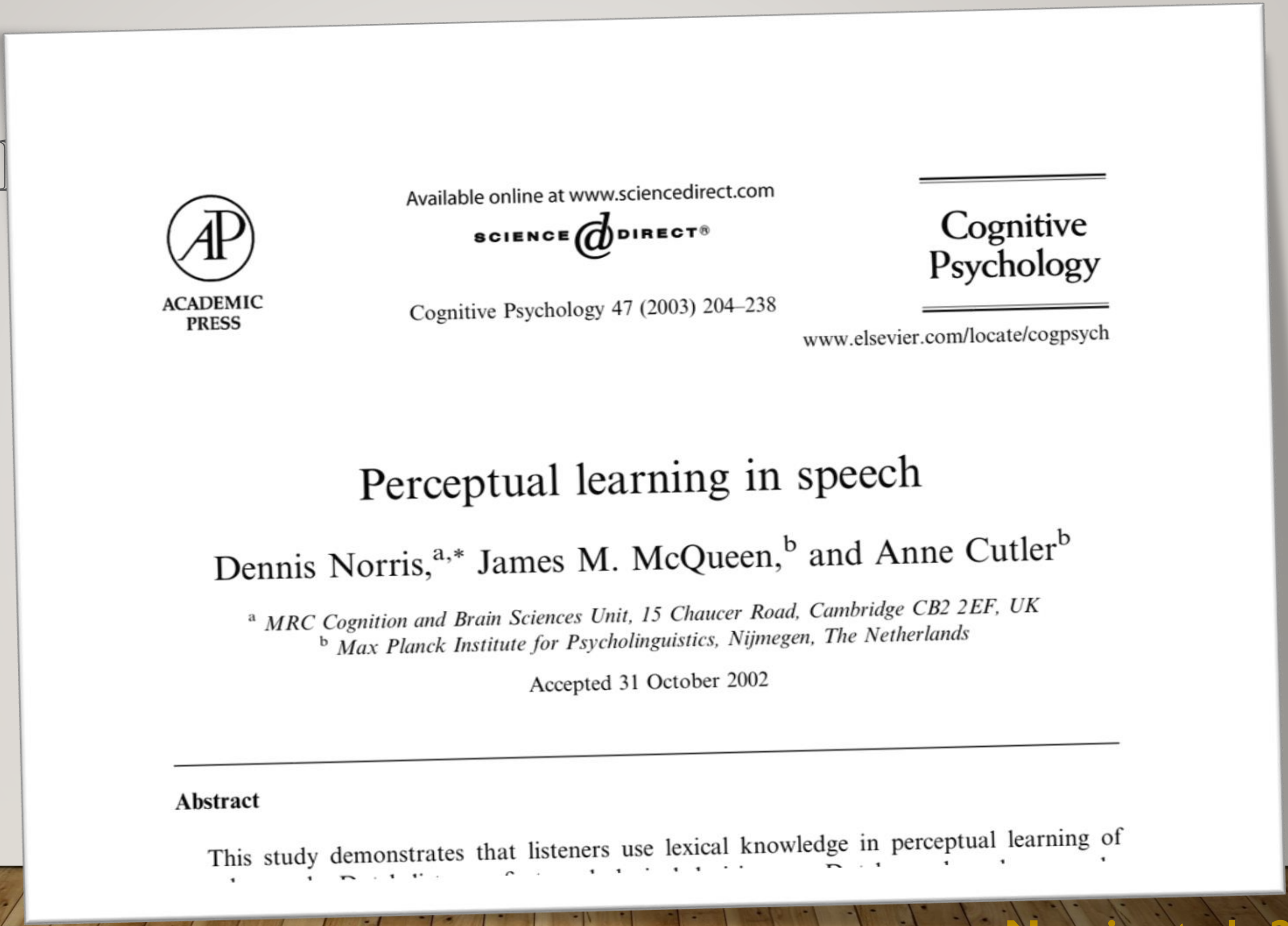
Reinisch, 2016; Ulusahin, Bosker, Meyer, & McQueen, *subm.*

Knowledge about the *usage* of prosody?

- Previous examples: learning about *average f*0 height/speaking tempo

- What about learning about how a given talker *uses* various suprasegmental cues to signal different prosodic categories?
  - Can we learn *how* Talker X happens to produce questions vs. statements?
  - Can we learn *which cues* Talker X likes to use to signal lexical stress?

Perce



ACADEMIC PRESS

Available online at www.sciencedirect.com

SCIENCE DIRECT®

Cognitive Psychology 47 (2003) 204–238

Cognitive Psychology

www.elsevier.com/locate/cogpsych

Perceptual learning in speech

Dennis Norris,[a,*] James M. McQueen,[b] and Anne Cutler[b]

[a] *MRC Cognition and Brain Sciences Unit, 15 Chaucer Road, Cambridge CB2 2EF, UK*
[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

Accepted 31 October 2002

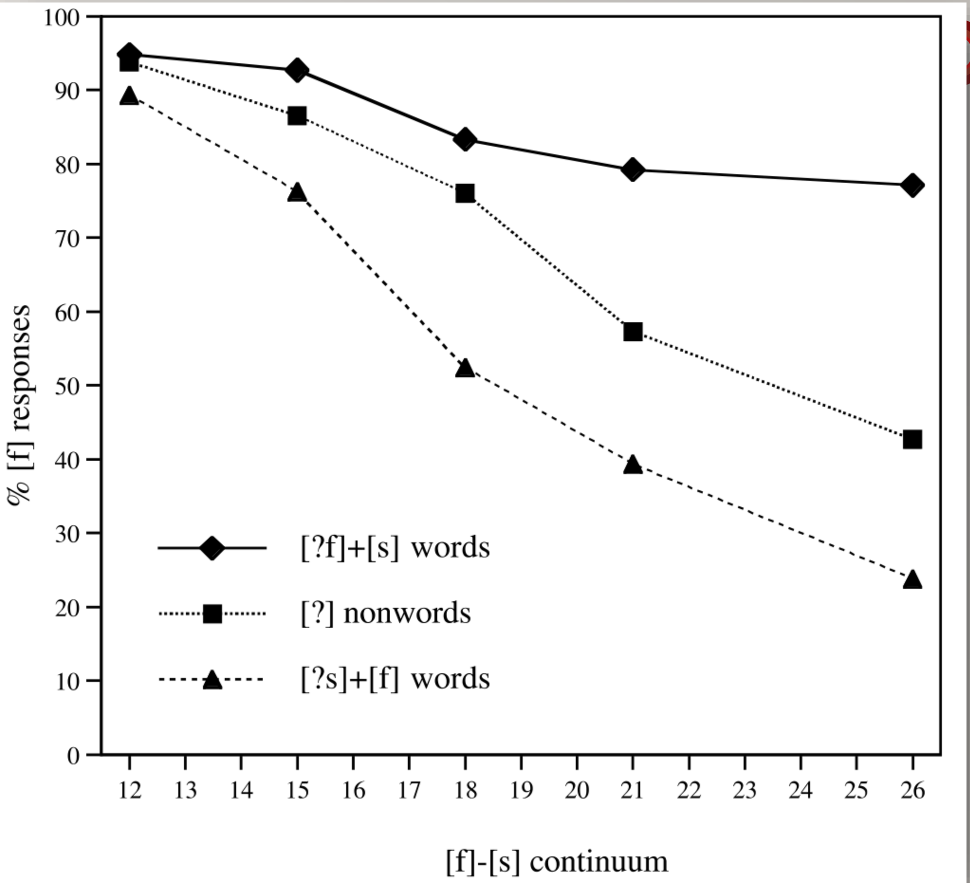
Abstract

This study demonstrates that listeners use lexical knowledge in perceptual learning of

Norris et al., 2003

Perceptual learning

- EXPOSURE:
  - Group 1: lexical decision t
  - Group 2: lexical decision t
  - Control group: lexical dec
- TEST:
  - All groups categorize the



Norris et al., 2003

Perceptual learning: segments

- EXPOSURE:
  - Group 1: lexical decision task: "platypu[?]", "giraffe", etc.
  - Group 2: lexical decision task: "platypus", "gira[?]", etc.
  - Control group: lexical decision: "dog", "cat", *ploo[?]
- TEST:
  - All groups categorize the same [s-f] continuum

➢ **Lexically-guided perceptual learning**

(a.k.a. phonetic retuning, recalibration, …)

Norris et al., 2003

Perceptual learning: segments

- **Lexically-guided perceptual learning…**
  - …generalizes to new words not encountered in exposure (e.g., [nai?]; McQueen et al., 2006)
  - …is talker-specific (no effect when testing a new talker; Eisner & McQueen, 2005)
  - …persists over time (12h; Eisner & McQueen, 2006)
  - …is largely phoneme-specific (learning about /d-t/ does not generalize to /b-p/; Kraljic & Samuel, 2006)
  - …is context-dependent (no learning when speaker has a pen in the mouth; Kraljic et al., 2008)

Perceptual learning: segments

- **Perceptual learning can be driven by a large range of sources**
  - Lexicon: platypu[?] = “platypus” (Norris et al., 2003)
  - Visual articulation: [?a] = “ba” with a video of a talker closing his lips (Bertelson et al., 2003)
  - Semantic context: “He cuts the loaf with a [nai?]” = “knife” (Jesse, 2021)
  - Contra-aural context: L [?a] + R [ba] = “ba” (Scott, 2020)
  - …

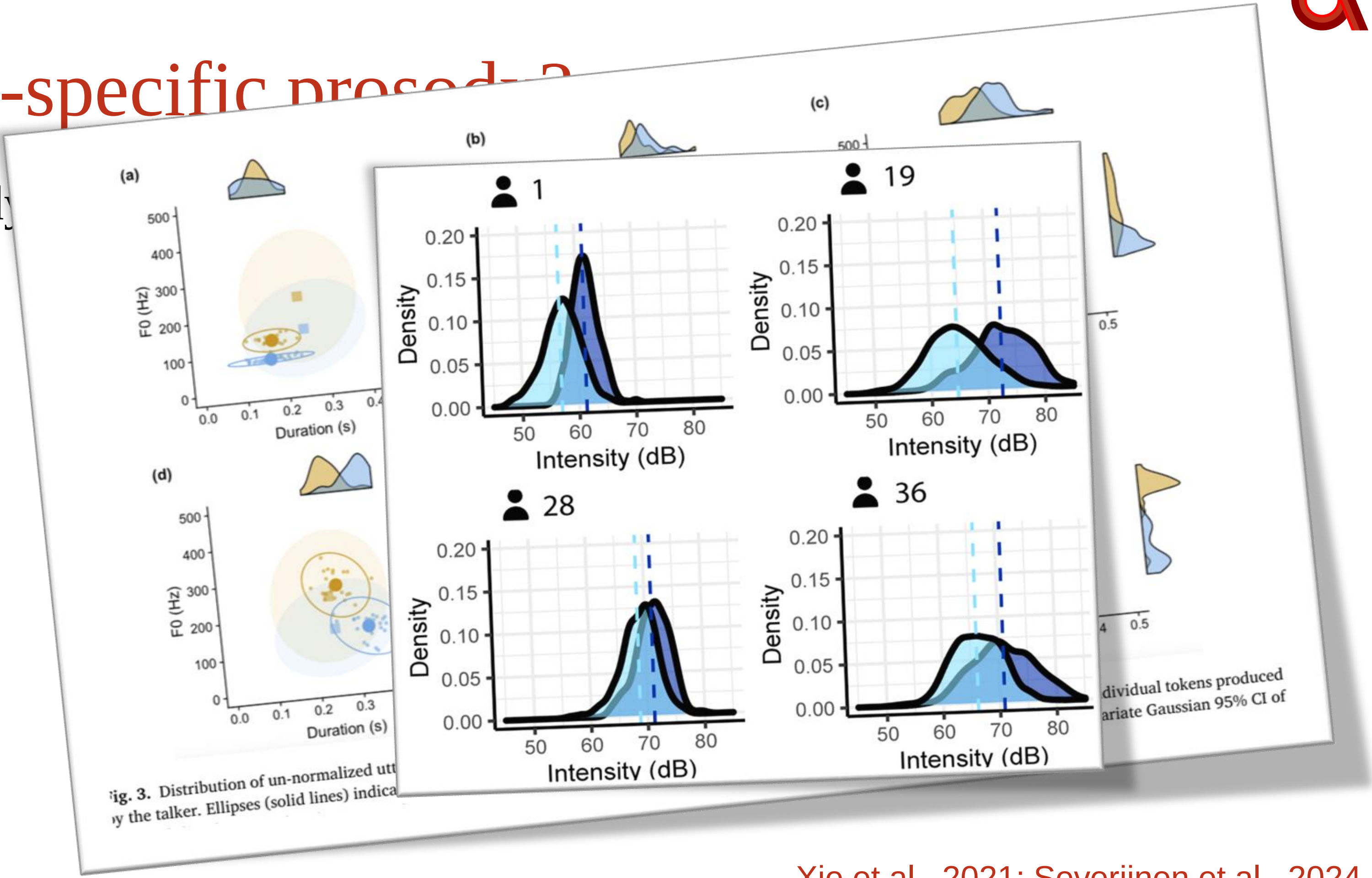
Perceptual learning: segments

- **Perceptual learning is useful!**
  - provides a perceptual mechanism to navigate the large variability in speech
  - allows listeners to track talker-specific pronunciation idiosyncrasies
    - Not just: "*On average,* this talker happens to produce overall longer VOTs"
    - But: "This talker happens to say /b/ a bit strangely" ~
      "this talker's category boundary between /b-p/ lies at a surprisingly high VOT"
  - is strongly related to how we 'tune into' foreign-accented speech

  - for reviews, see Kleinschmidt & Jaeger, 2015; Samuel & Kraljic, 2009.

Talker-specific prosody?

- Prosody



ïg. 3. Distribution of un-normalized utt ... dividual tokens produced
ɔy the talker. Ellipses (solid lines) indica ... ariate Gaussian 95% CI of

Xie et al., 2021; Severijnen et al., 2024

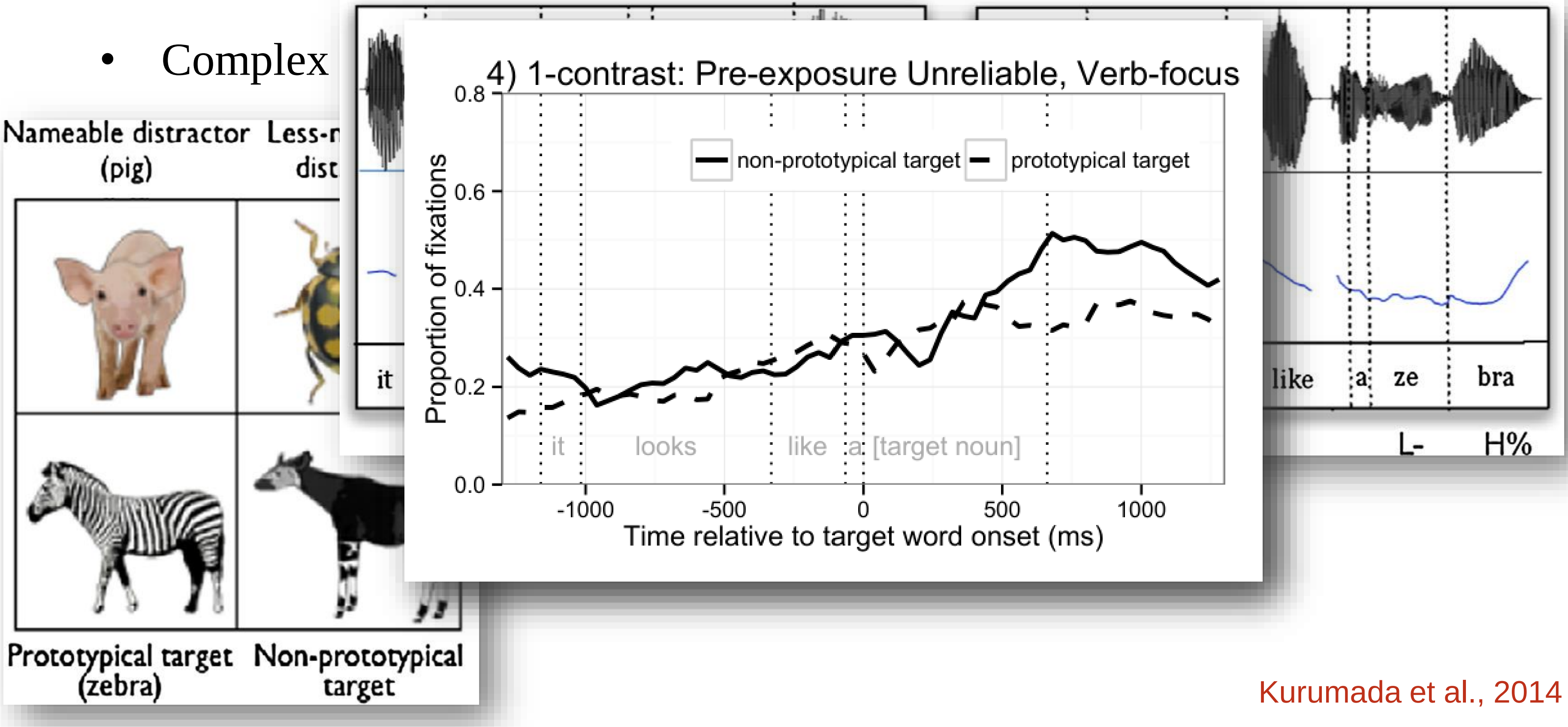
Talker-specific prosody?

- Prosody is also produced in talker-dependent manner

- Do people also learn about talker-specific prosody?
  - prosody less commonly distinguishes between words
    - hence, less feedback to the listener through lexical disambiguation
  - the same prosodic cues can convey multiple types of prosody
    - more complex mapping between acoustic input and perceptual categories
  - not all types of prosody are equally crucial for speech perception
    - lexical stress only lexically distinctive in some words, in some languages
    - should listeners spend cognitive resources on perceptual learning about lexical stress?

Learning about talker-specific prosody

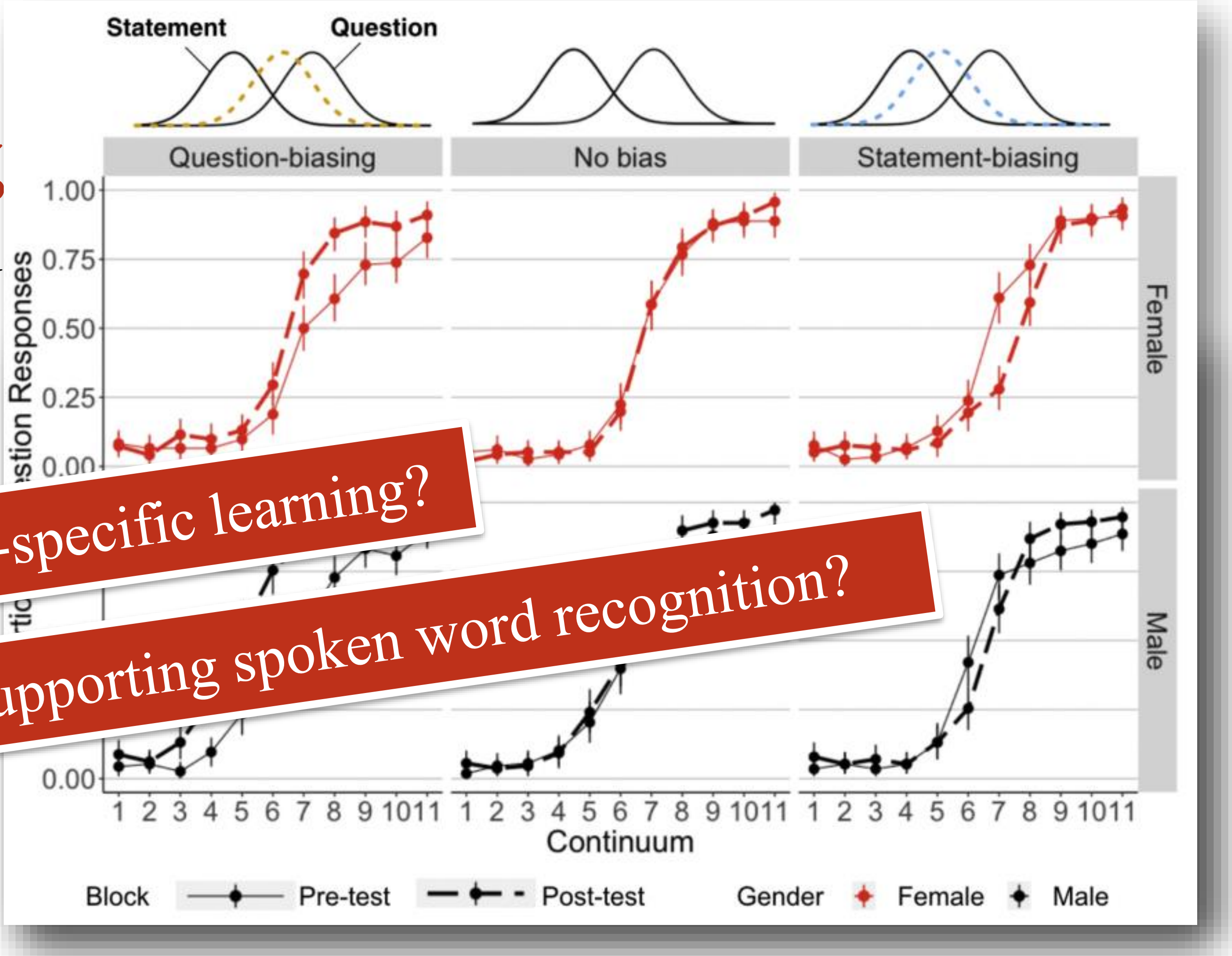
- Complex



Kurumada et al., 2014

Learning

- Question v

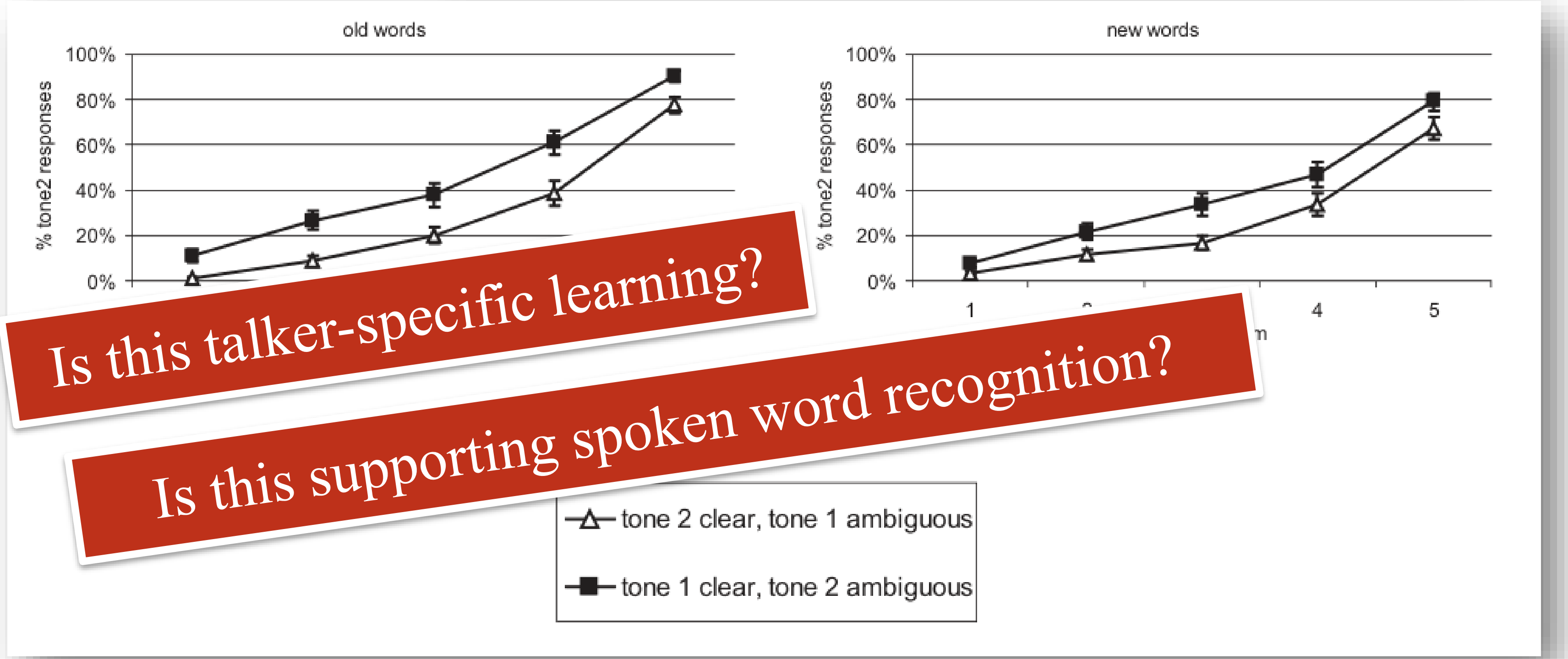


Is this talker-specific learning?

Is this supporting spoken word recognition?

Xie et al., 2021

Learning about talker-specific prosody



Is this talker-specific learning?

Is this supporting spoken word recognition?

Mitterer et al., 2011

Learning about talker-specific prosody

- Lexical stress

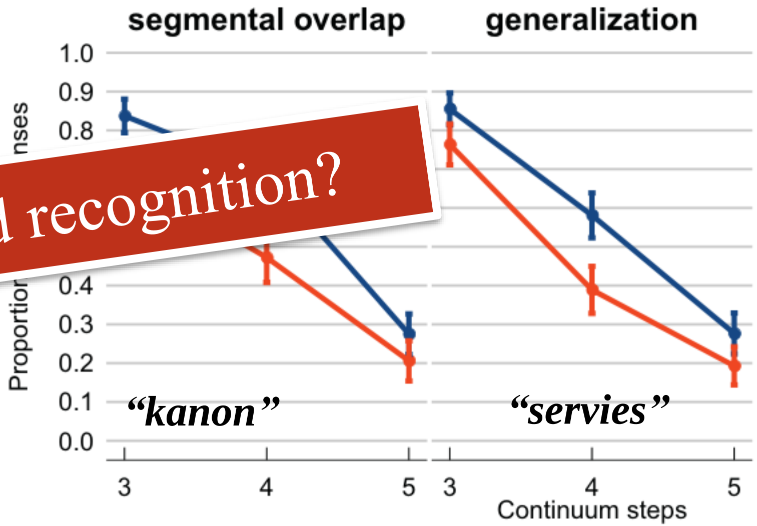
EXPOSURE (AV)

- Group 1: /ka.n /?
  aNON
- Group 2: /ka.n /?
  non

Is this talker-specific learning?

Is this supporting spoken word recognition?

TEST (A-only)



Bosker (2022, *Language and Speech*)

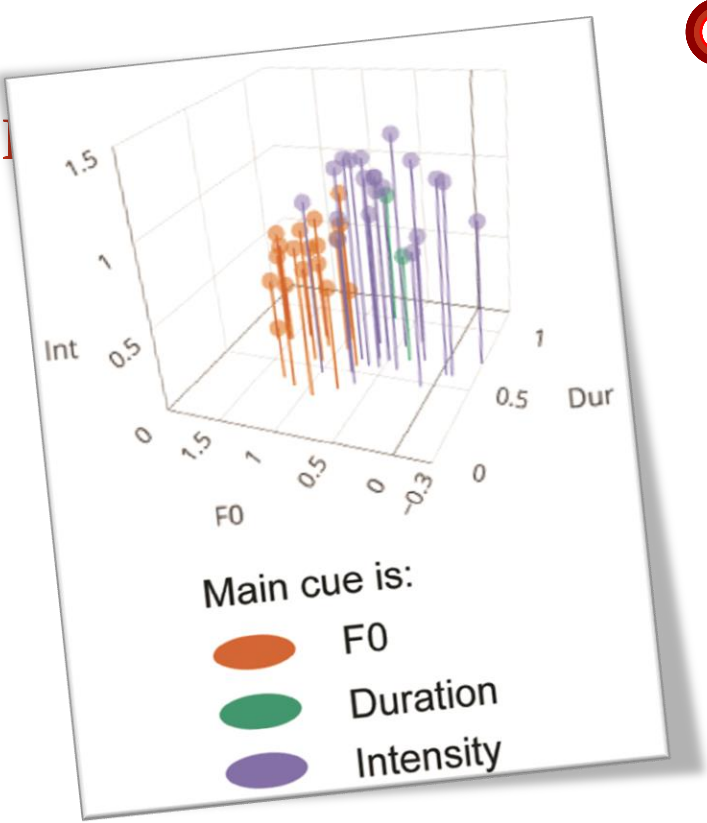
Learning about cue-weighting?

- So far: people adjust their perception of prosody (in a talker-specific manner?) when exposed to:
    - An unreliable talker (weakening the mapping between prosody and referent)
    - An ambiguous talker (shifting the category boundary)

- Do people also adjust to talker-specific cue-weights?

Learning about cue-weighti

- Remember these?



Severijnen, Bosker, & McQueen, 2021; 2024

Learning about cue-weighting?

- People have unique cue-weights of lexical stress cues
- …yet the variability is not unbounded
- Can listeners learn that X is an *‘f0-user’* but Y is an *‘intensity-user’*?

Severijnen, Bosker, et al., 2021; 2023; 2024

Learning about cue-weighting?

EXPOSURE PHASE

TEST PHASE

slow down in RTs for

lker-incongruent cues

Is this talker-specific learning?

Is this supporting spoken word recognition?

Group 1: talke

pectedly

uses *intensity*

Group 2: talker only uses intensity

talker unexpectedly uses *f0*

Severijnen, Bosker, et al., 2021; 2023; 2024

Learning about cue-weighting?

EXPOSURE PHASE

TEST PHASE

when in doubt,

isteners go for talker-congruent cue

...even in d ~25min delay)

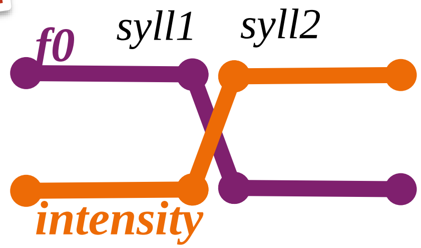
Group 1: talker ... syllable!

Group 2: talker ... stress on 2nd syllable!

Is this talker-specific learning?

Is this supporting spoken word recognition?

Will this generalize?



Severijnen, Bosker, et al., 2021; 2023; 2024

Wrap-up of today

- Vast acoustic variability in how prosody is produced.
    - Group-level differences
    - Talker-specific idiosyncrasies
- Listeners can adjust their perception in a talker-specific fashion
    - …of segments, but also prosody
        - …in order to cope with unreliable prosody-referent mappings;
        - …ambiguous cues to prosodic categories;
        - …talker-specific cue-weights
- Learning supports and speeds up perception and spoken word recognition

Wrap-up of today

- Open questions:
    - When does learning arise? (exposure: how much/how long/how (un)reliable?)
    - When does learning fail? (test: how long do effects persist/unlearning?)
    - When do people *not* learn?
    - When is learning talker-specific? When does it generalize over talkers?
    - Does it generalize to new words/utterances?
    - Impact/effect size in real-life communication?

Next up:

- Lecture 5: *Audiovisual integration of multisensory prosody*

Hans Rutger Bosker

Speech Perception in Audiovisual Communication [SPEAC] lab

Donders Institute, Radboud University, Nijmegen, The Netherlands

https://hrbosker.github.io

hansrutger.bosker@donders.ru.nl

European Research Council

Established by the European Commission