Prosody in Speech Perception

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

Who am I?

- Assistant Professor at Psychology, Social Sciences, Radboud University
- Principal Investigator of the SPEAC lab, Donders Institute, RU

- 2005 – 2010: BA & MA in Linguistics, Leiden (!)
- 2010 – 2013: PhD in Psycholinguistics, Utrecht
  with Nivja de Jong, Hugo Quené, and Ted Sanders
- 2014 – 2023: Postdocs at RU and Max Planck Institute, Nijmegen
- ERC Starting Grant in 2022

What do I do?

speech
language
perception
rate
prosody
stress
psychology
evidence
acoustic
lexical
native
fluency
tracking
normalization
attention
research
human
society
experimental
cognition
processing
gestures
open
cues
non-native
influences
cognitive
neuroscience
contextual
discourse
disfluency
audiovisual
spontaneous
listening
data
word
dutch
behavior
nonnative
access
linguistic
psychophysics
modulations
sound
influence
workshop

Who are you?

- Large majority is RMA students, some PhDs
- Leiden and Groningen in equal first place
    - No Tilburg, Maastricht, Rotterdam sign-ups?! ☺
- Backgrounds
    - Phonology / Phonetics
    - Syntax / Semantics
    - Discourse / Communication

- Take away: **raise your hand if a concept or term is unfamiliar to you!**

What's this course about?

- **Prosody in Speech Perception**
- Course aims:
  - to be familiar with key concepts in the area of speech prosody and speech perception
  - to be familiar with recent advances and new paradigms in the speech perception literature
  - to understand how prosody influences the perception of vowels, consonants, and words
  - to understand the different processing mechanisms that underlie these influences
  - to understand the open issues and debates in the field of speech perception

Practicalities

- Course description:
  - https://hrbosker.github.io/resources/course-materials/prosody-in-speech-perception
  - All descriptions, materials, PDFs, links, and slides can be found here

- Slides are made available after each lecture.

Expectations

- Interactive: raise your hand, speak up, ask questions, email!
- Be prepared: read the recommended literature!
- Be there!

Any questions?

What is this thing called ‘prosody’?

What is ‘prosody’?

- Nooteboom et al. (1978); Rietveld & Van Heuven (2009, p275)
  - all speech aspects that cannot be traced back to the vowels and consonants
  - ~ *suprasegmental* (Lehiste, 1970)
- Hayward (2000, p273): “patterned variation in pitch, force, and duration”
- Arvaniti (2009, p1): “Prosody is an umbrella term used to cover a variety of interconnected and interacting phenomena, namely stress, rhythm, phrasing, and intonation. The phonetic expression of prosody relies on a number of parameters, including duration, amplitude, and fundamental frequency (*f*0).”
- Beckman & Edwards (1994, p8): “prosody is the organizational framework that measures off chunks of speech” ~ metrical theory (Liberman, Pierrehumbert)

Which phenomena are ‘prosodic’?

Some examples perhaps:

- Intonation?
- Lexical stress?
- Sentence accent?
- Question vs. statement?
- Pausing, chunking?
- Emotion?
- Register?

Which phenomena are **not** ‘prosodic’?

But what about:

- Vowel length?
- Final lengthening?
- Breathing? Domain-initial strengthening?
- Creaky voicing?
- Speech rate?
- Speech reductions?
- Lexical tone?
- Disfluencies?
- Motherese?
- Clear speech, Lombard speech?
- Reverberation from room acoustics
- Visual signals, such as lip movements, hand gestures, facial expressions?

Which acoustic cues are ‘prosodic’?

Some examples perhaps:

- Fundamental frequency (*f*0)?
- Intensity?
- Duration?

But what about:

- Formants?
- Spectral tilt?
- Vowel quality?
- Room acoustics?

In this course…

- I will adopt a ‘broad’ definition of prosody: any speech phenomenon, in any modality, that cannot be traced back to the vowels and consonants in speech
- Suprasegmentals: *f*0, intensity, duration
- Segments: vowels and consonants

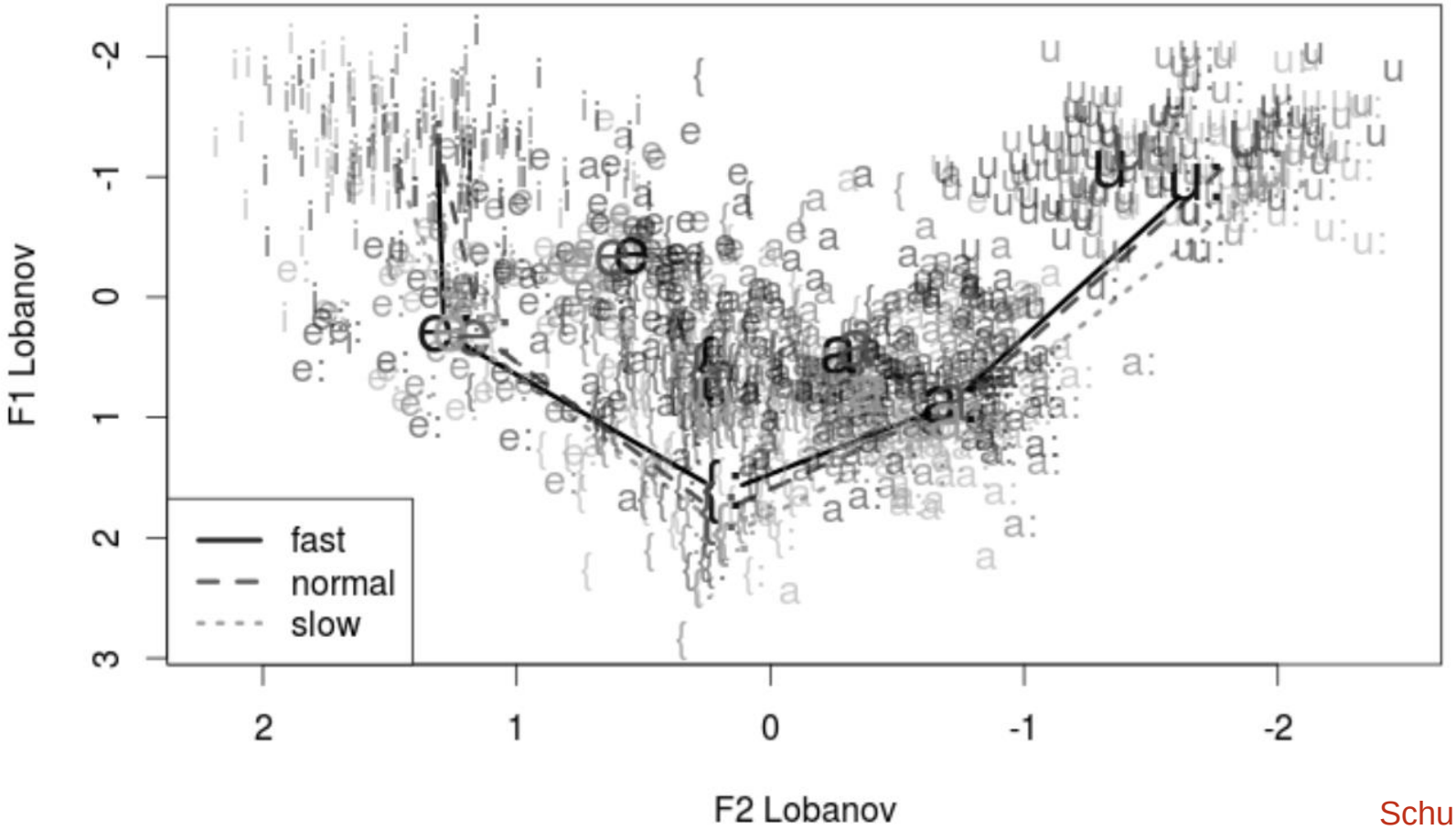
In this course…

- I will try to convince you that prosody can change which words you hear!
  - …lexical stress, speech rate, rhythm, speaker's vocal tract size, room acoustics, talker-specific pronunciations, simple up-and-down hand gestures
- 'Suprasegmental' influences on 'segmental' perception, emphasizing the problematic nature of this distinction
  - cf. Eisner & McQueen (2018); McQueen & Dilley (2021)

In this course…

- Core premise: speech perception is hard!
- ‘Lack of invariance’ problem:
    - the same phoneme can be produced in a zillion different ways

Vowels produced by a single native speaker of Finnish



Schulz et al., 2016

In this course…

- Core premise: speech perception is hard!
- ‘Lack of invariance’ problem:
  - the same phoneme can be produced in a zillion different ways
    - one-to-many mapping of phoneme > audio
  - the same acoustic recording can be perceived as word A by some, but as word B by others.
    - one-to-many mapping of audio > phoneme


In this course…

- Core premise: speech perception is hard!
- ‘Lack of invariance’ problem:
    - the same phoneme can be produced in a zillion different ways
        - one-to-many mapping of phoneme > audio
    - the same acoustic recording can be perceived as word A by some, but as word B by others.
        - one-to-many mapping of audio > phoneme
- Many-to-many mappings between phonetics > phonology
- Prosody to the rescue!

In this course…

- Five mechanisms by which prosody influences segmental perception
  - general-auditory normalization for prosody
  - neural tracking of prosody
  - prosody-guided prediction
  - talker-specific learning of prosody
  - audiovisual integration of multisensory prosody

- Through these combined mechanisms, prosody supports speech perception, thus overcoming the large variability in speech.

Any questions?

Preparatory reading:

- Arvaniti, A. (2020). The Phonetics of Prosody. In S. Calhoun (Ed.), *Oxford Research Encyclopedia of Linguistics*. Oxford: Oxford University Press. doi:10.1093/acrefore/9780199384655.013.411.
- Nooteboom, S., Brokx, J. P. L., & De Rooij, J. J. (1978). Contributions of Prosody to Speech Perception. In W. J. M. Levelt and G. B. Flores d'Arcais (Eds.), *Studies in the Perception of Language*. p.75-107. New York: Wiley. Open fulltext.

Lecture 1: *normalization*

Bosker, H. R., Sjerps, M. J., & Reinisch, E. (2020). Temporal contrast effects in human speech perception are immune to selective attention. *Scientific Reports,* 10: 5607. doi:10.1038/s41598-020-62613-8.

Hans Rutger Bosker

Speech Perception in Audiovisual Communication [SPEAC] lab

Donders Institute, Radboud University, Nijmegen, The Netherlands

https://hrbosker.github.io

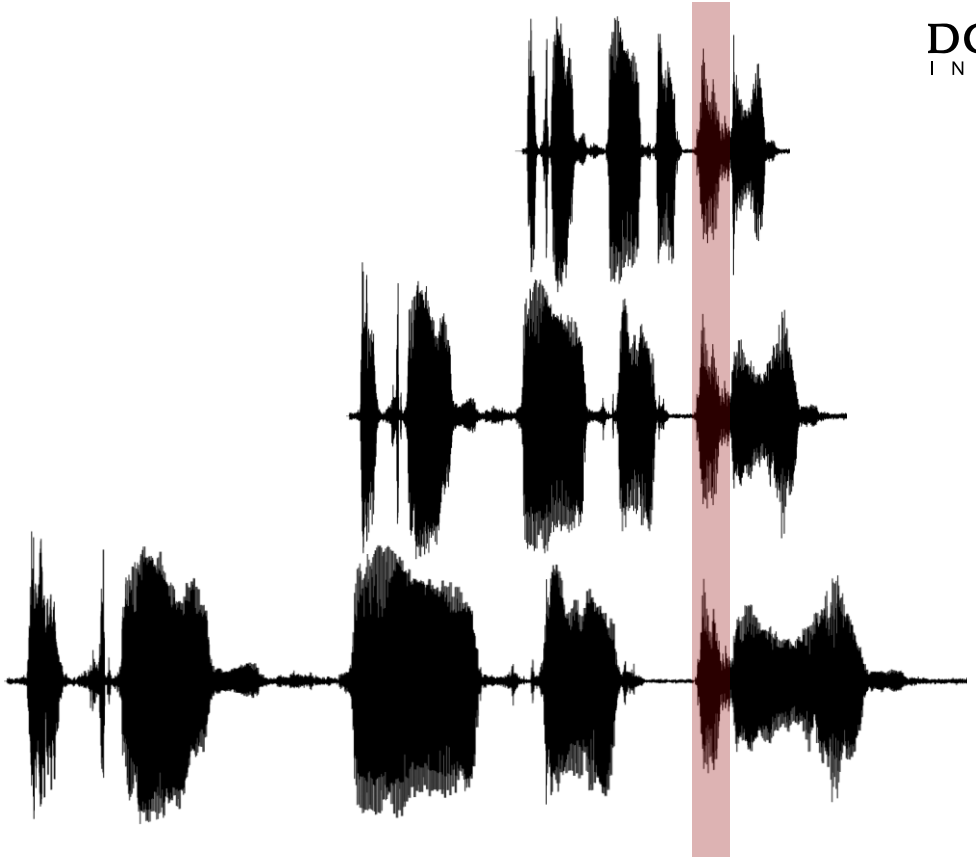
hansrutger.bosker@donders.ru.nl

DONDERS INSTITUTE

FAST

ORIGINAL

SLOW



That's one small step... *for (a)* *man*

SPEAC

cf. Baese-Berk et al. (2016)

High F1 in context

+ /u/

+ /o/

Low F1 in context

example stimuli drawn from Sjerps et al. (2019)

Acoustic context effects

- We don’t perceive absolute time or frequencies
- Perception depends on the (here: acoustic) context!

for a review, see Stilp, 2019, *WIREs*

Acoustic context effects

- Rate normalization
  - perception of a target duration depends on the surrounding speech rate
  - target sounds relatively long if embedded in a fast context
    …but as relatively slow if embedded in a slow context
- Spectral normalization
  - perception of a target frequency depends on the surrounding spectral context
  - target sounds relatively high-pitched if embedded in a low-pitched context
    …but as relatively low-pitched if embedded in a high-pitched context

➢ Note: both effects are contrastive in direction (slow context? fast target!)

SPEAC

for a review, see Stilp, 2019, WIREs

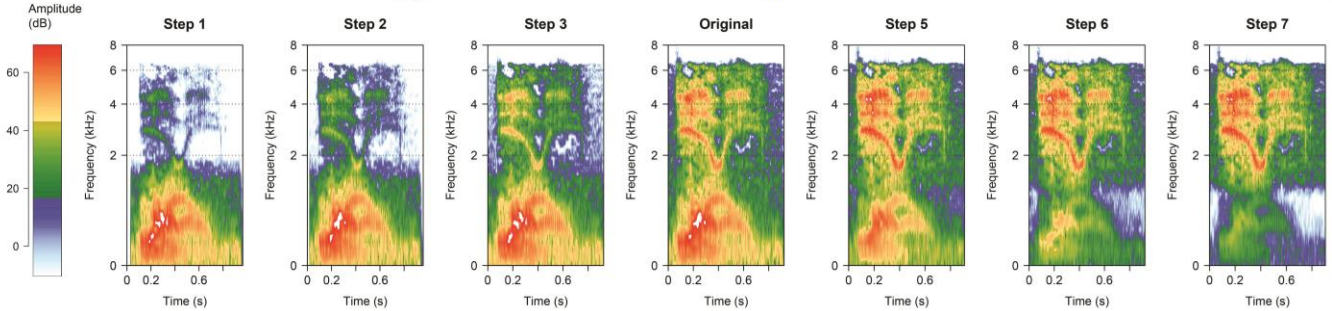
Spectral normalization

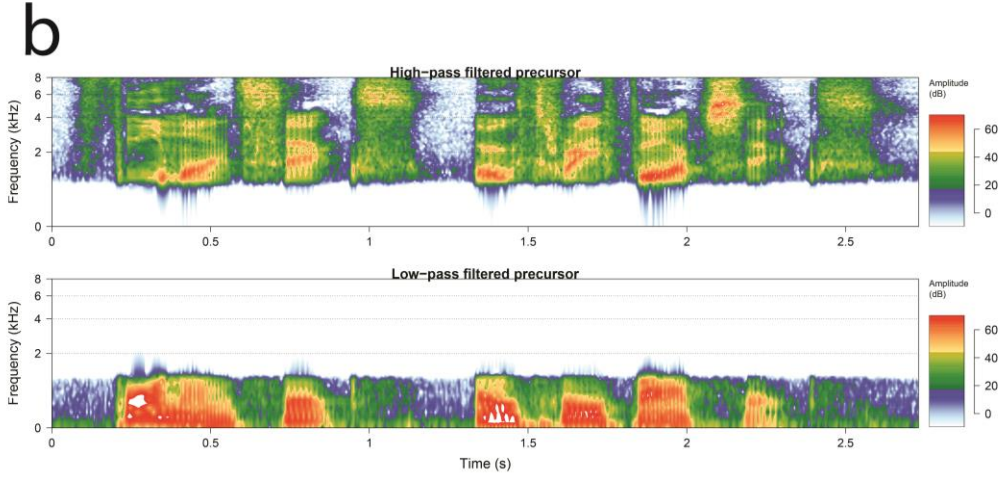
- The spectral properties of the surrounding context contrastively influence the perception of spectral properties of a given target sound
  - Does the context have high {formants, *f*0, power in frequency band *b*}?
  - Then the target sound will be perceived as having relatively low { … }!
- Spectral normalization demonstrated with vowel perception (formants), lexical tone height (*f*0), lexical items (frequency bands), etc.

Ladefoged & Broadbent, 1957; Moore & Jongman, 1997; Sjerps et al., 2011; Stilp & Assgari, 2018; Stilp, 2019; Watkins, 1991

Hans Rutger Bosker: JASA Express Letters https://doi.org/10.1121/1.5070144 Published Online 5 December 2018

Putting Laurel and Yanny in context



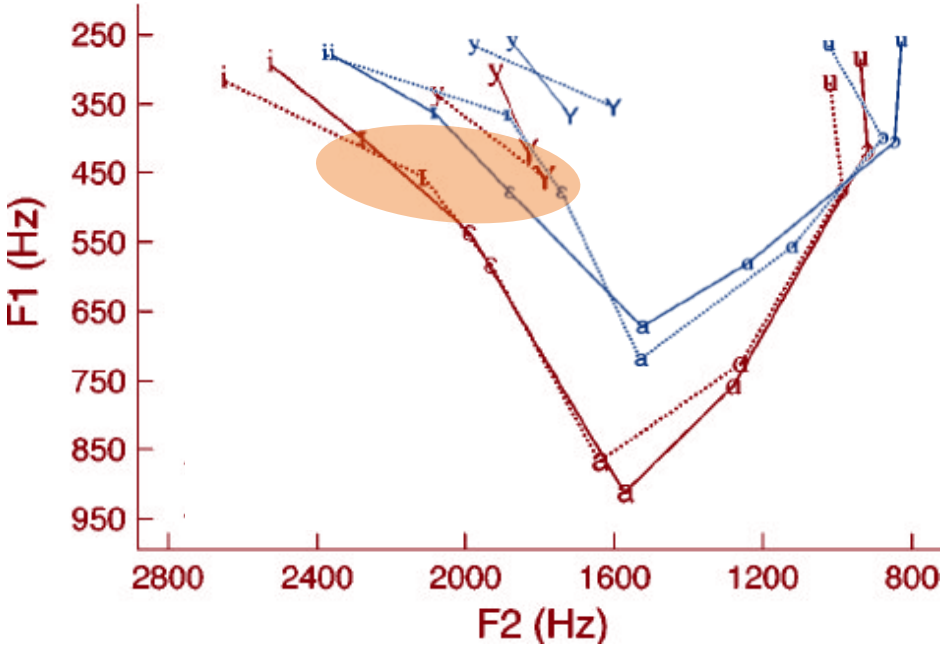


Is this prosody?

Bosker, 2018, *Journal of the Acoustical Society of America*

Spectral normalization: **why?**

- Helps overcome variability in
  - vocal tract size
  - talker pitch
  - room acoustics



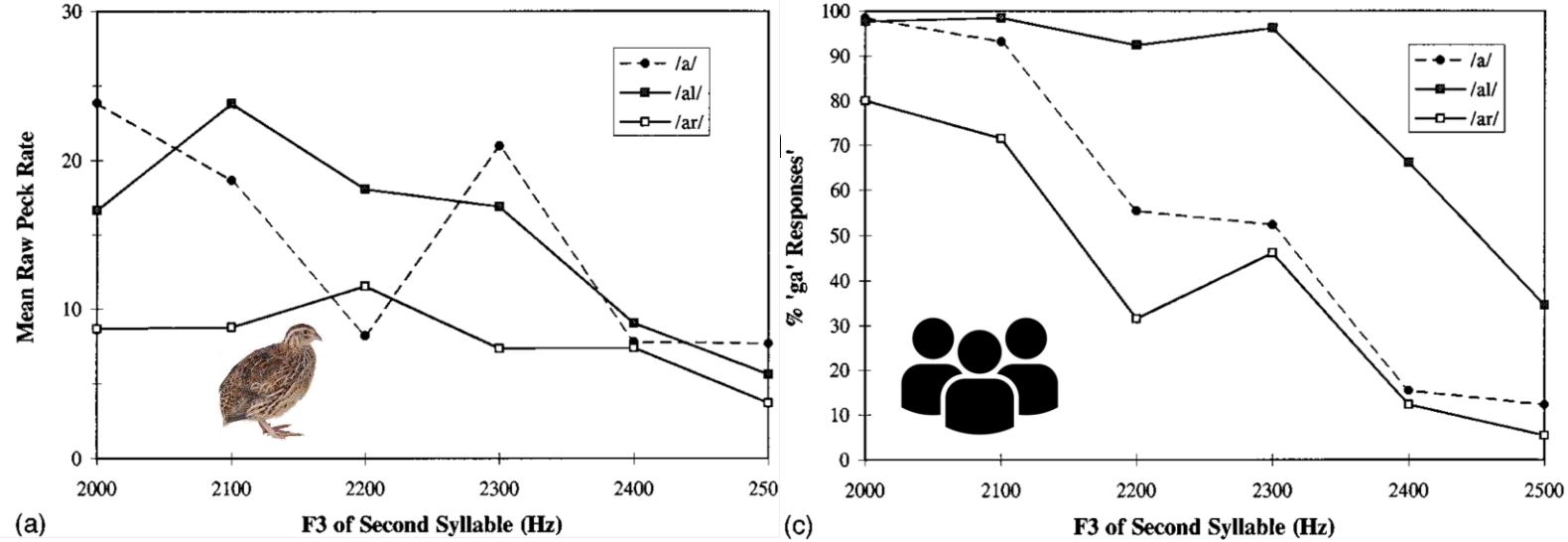
Adank, Van Hout, & Smits, 2004

Spectral normalization: **how?**

- How does spectral normalization arise mechanistically?
  - Does it need a human?
  - Does it need speech?
  - Does it need a brain?
  - Does it require attention?

Spectral normalization: **how?**

- Does it need a human?



Lotto, Kluender, and Holt, 1997

Spectral normalization: **how?**

- Does it need speech?
  - No.

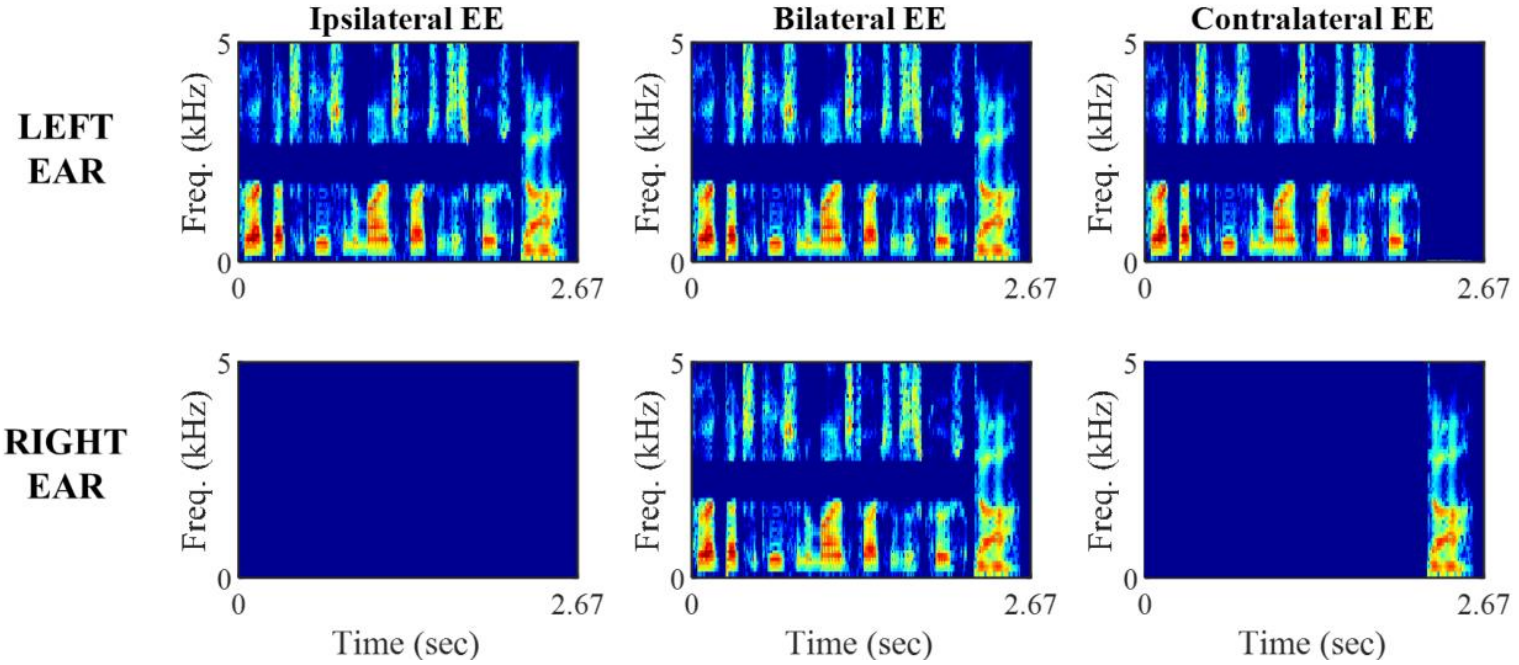
  - Signal-correlated noise (Watkins, 1991; Stilp 2021)…
    …pure tone sequences (beep trains: Holt, 2005; 2006)…
    …musical instruments (Stilp, Alexander, Kiefte, & Kluender, 2010; Lanning & Stilp, 2020)
    all induce spectral normalization

Spectral normalization: **how?**

- Does it need a brain?
  - No…

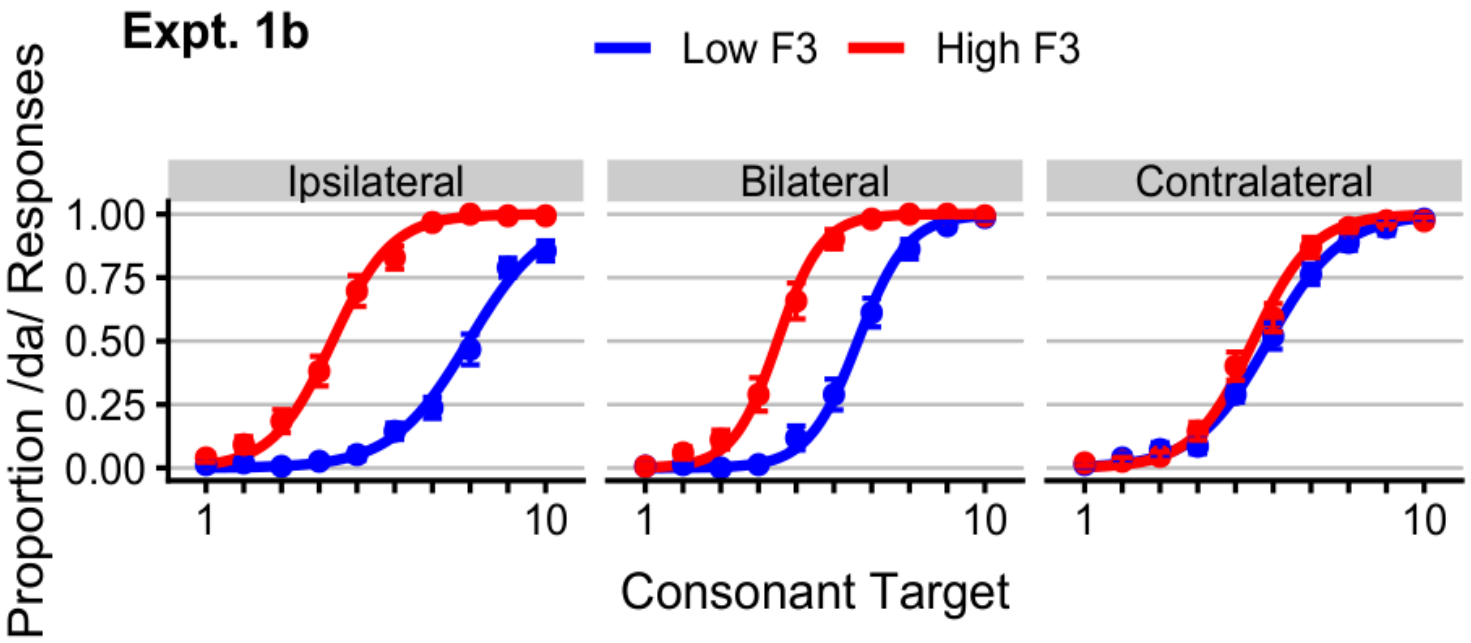
Spectral normalization: **how?**

- Does it need a brain?
  -



Stilp, 2020

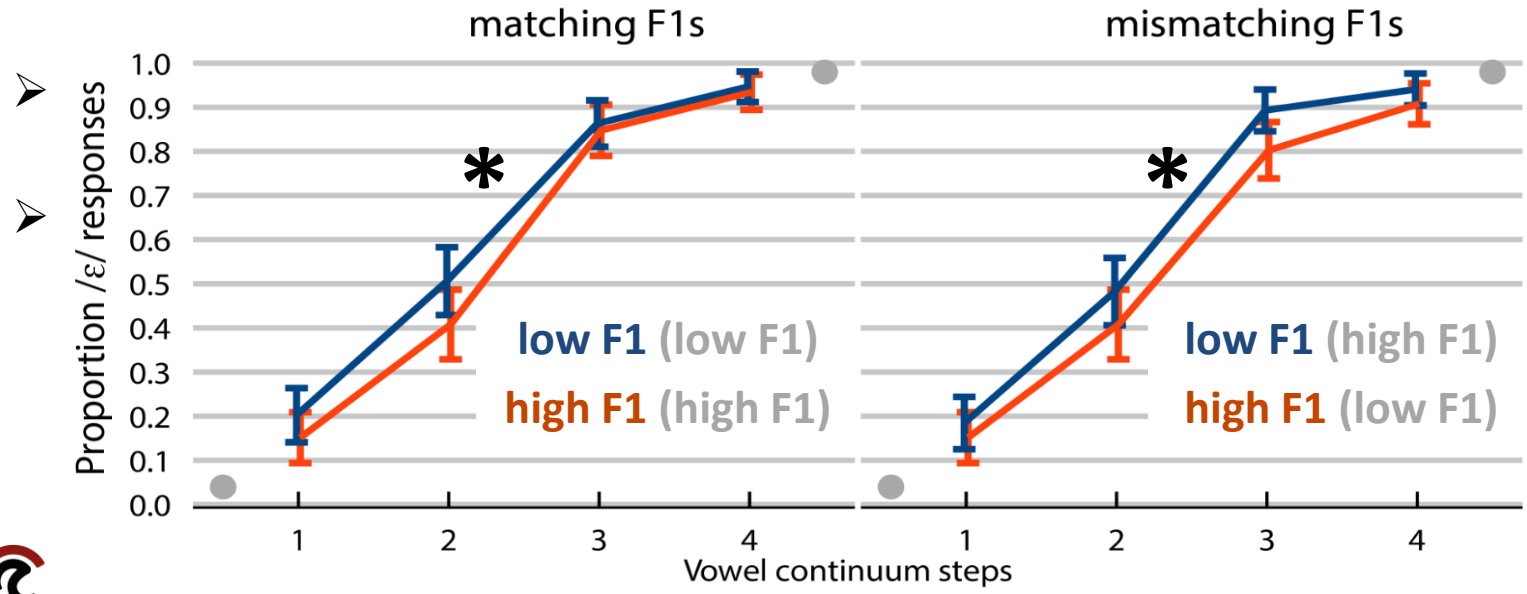
Spectral normalization: **how?**

- Does it need a brain?
  - 

Stilp, 2020

Spectral normalization: **how?**

- Does it need a brain?
  - Experiment 2a

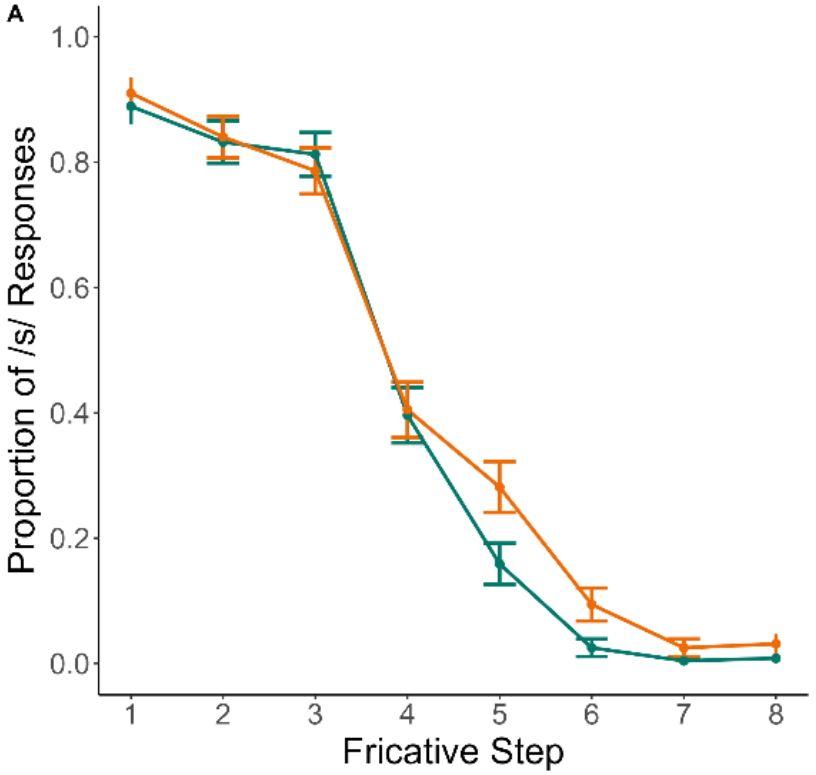


Bosker et al., 2020, *AP&P*

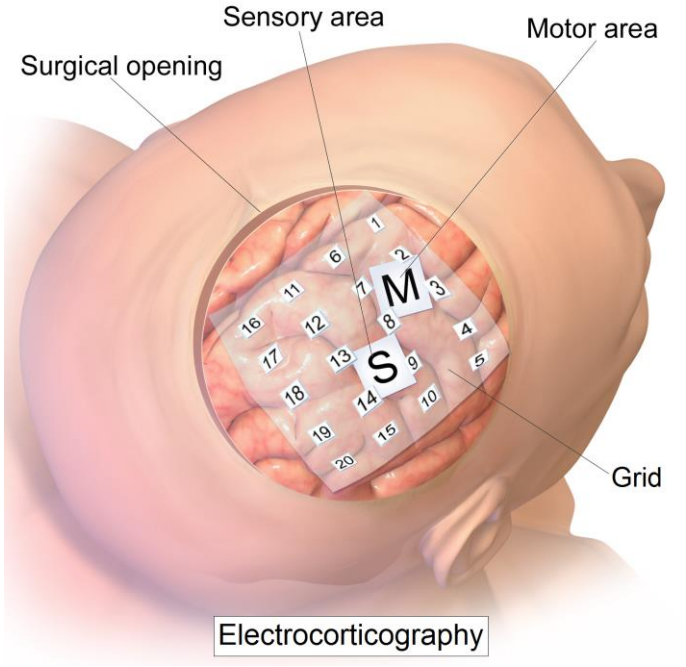
Spectral normalization: **how?**

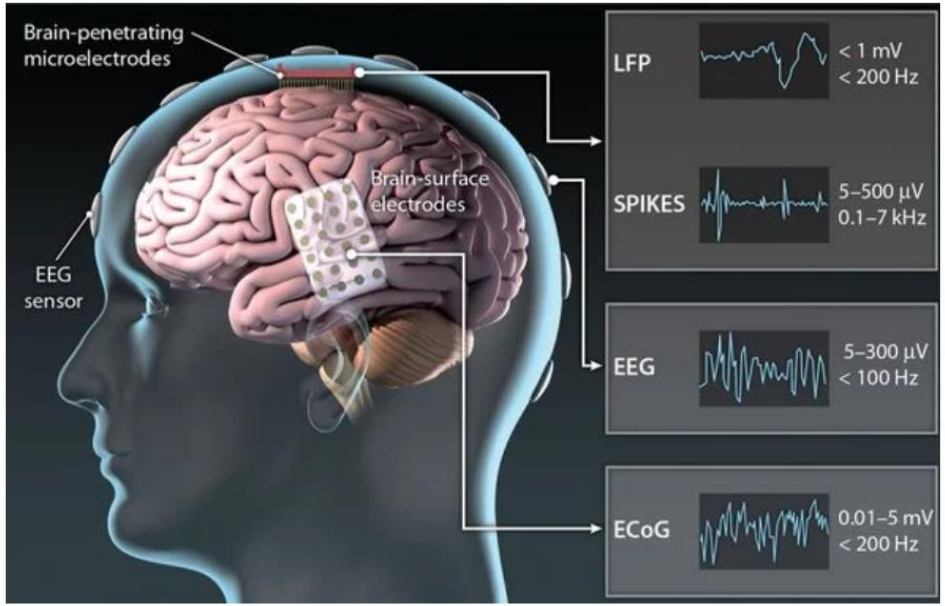
- Does it need a brain?
  - No… and yes?

  - ➢ Sentences from 200 talkers induce sentence from a single talker (Assga
  - ➢ Selectively attending to one of two normalization in the direction of th Feng & Oxenham, 2018)
  - ➢ Previously acquired knowledge ab (Ulusahin et al., *submitted*)
  - ➢ Visual influences? (*tbc* on Day 5)

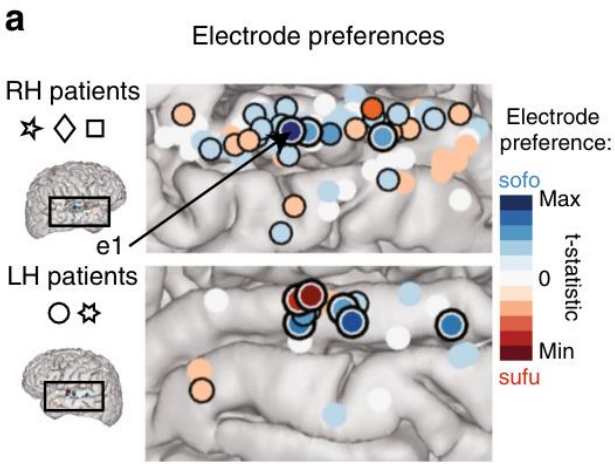


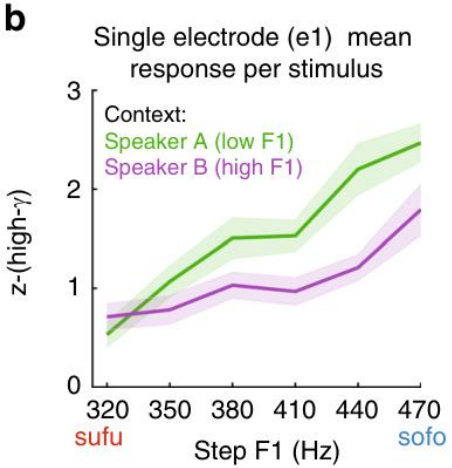
Spectral normalization: **neurobiology?**

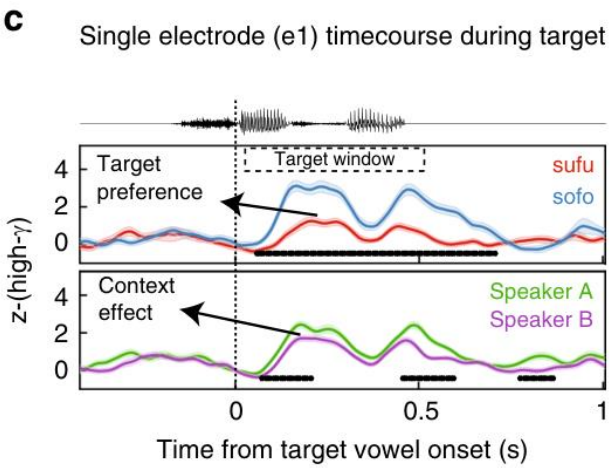




Spectral normalization: **neurobiology?**

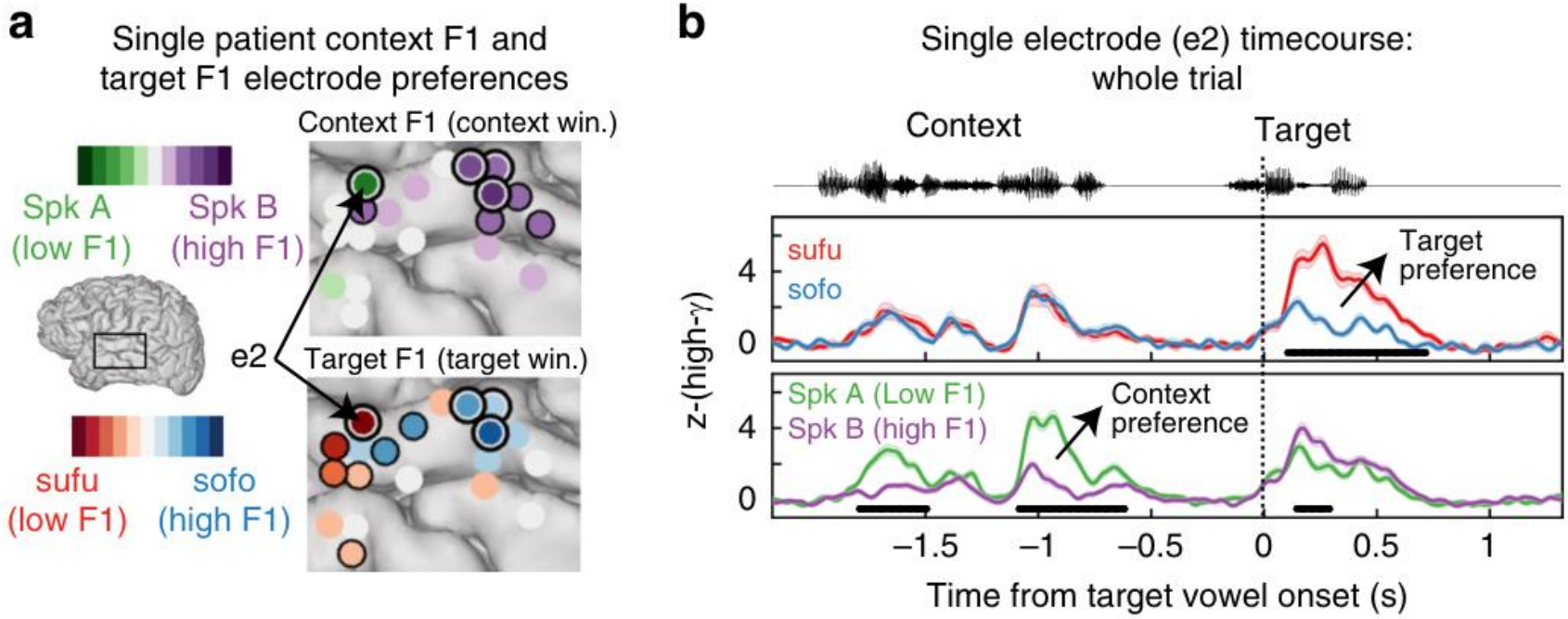






Sjerps et al., 2019, *Nature Communications*

Spectral normalization: **neurobiology?**



Sjerps et al., 2019, *Nature Communications*

Spectral normalization: **neurobiology?**

- Normalized representations of vowels in parabelt auditory cortex (STG)
- 500 ms between context and target; unlikely to be inherited from more peripheral regions
- General auditory contrast enhancement model of normalization (phoneme invariant)
- Sensory adaptation: neural fatigue & (inhibitory) interactions between separate populations of neurons

Sjerps et al., 2019, *Nature Communications*

Spectral normalization: **summary**

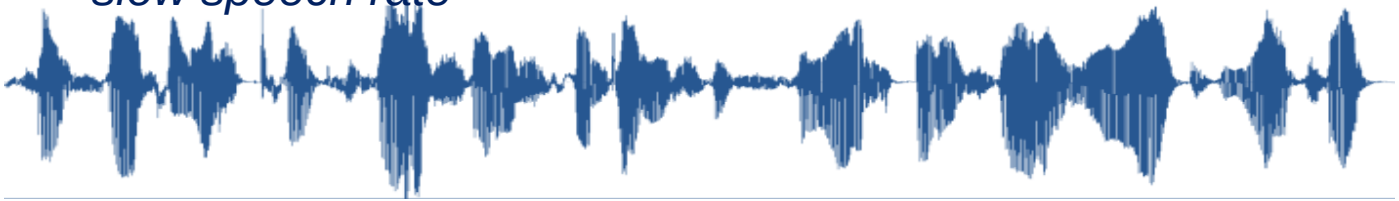
- We perceive target speech relative to the spectral properties of the surrounding acoustic context
- Spectral normalization seems to involve general auditory processing levels (i.e., not speech specific), as it is observed in nonhumans, can be induced by nonspeech, and does not (always) require central processing
- Neurobiological evidence points towards early auditory cortex (STG).
- Increasing evidence for more higher-level cognitive influences on spectral normalization.

SPEAC

for a review, see Stilp, 2019, WIREs

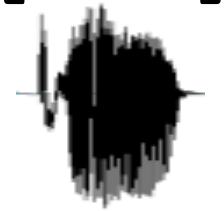
Rate normalization

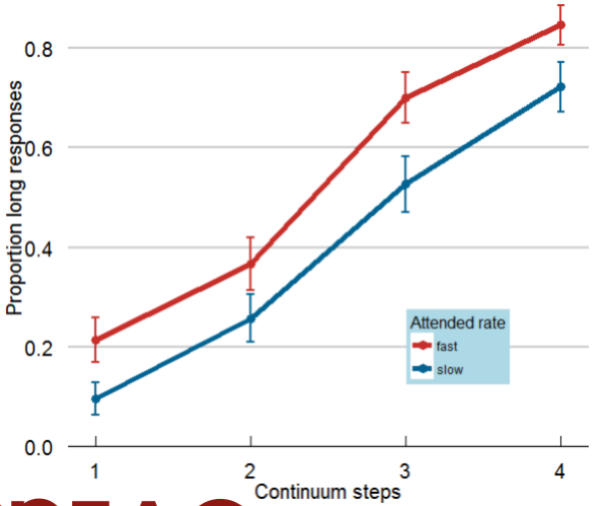
slow speech rate

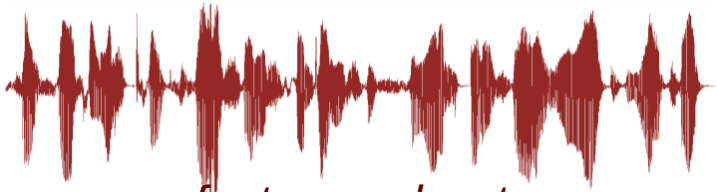


“tear”

[tɛɹ:] ?







fast speech rate

“terror”

for a review, see Stilp, 2019, WIREs

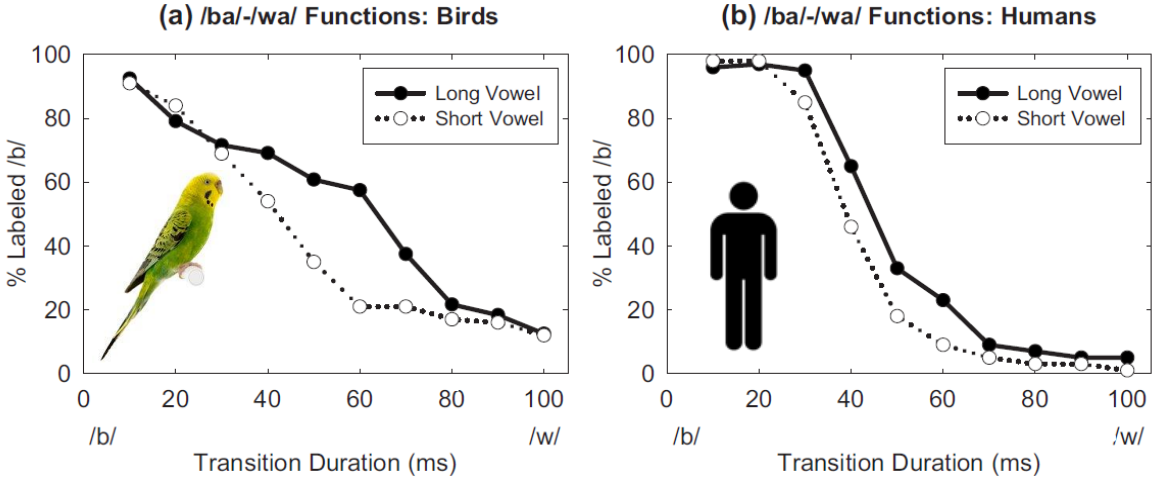
Rate normalization: **why?**

- Massive variability in speech rate
- Speech rate variation can make a short /ɑ/ in Dutch have the same duration as a long /a:/ (e.g., *tak* “branch” vs. *taak* “task”)
- Normalization for surrounding speech rate overcomes part of this challenge

➢ Observed for vowel length (Bosker et al., 2017), voice onset time (/b-p/; Miller & Liberman, 1979), formant transition durations (/b-w/; Wade & Holt, 2005), lexical stress (Reinisch et al., 2011), word segmentation (topic vs. top pick; Pickett & Decker, 1960), reduction (*for (a) man*; Dilley & Pitt, 2010; Baese-Berk et al., 2016), etc…

Rate normalization: **how?**

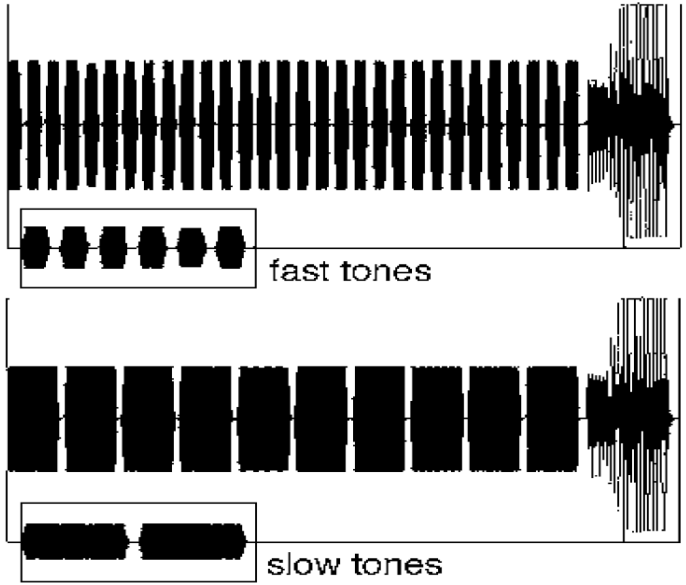
- Doesn't need humans!

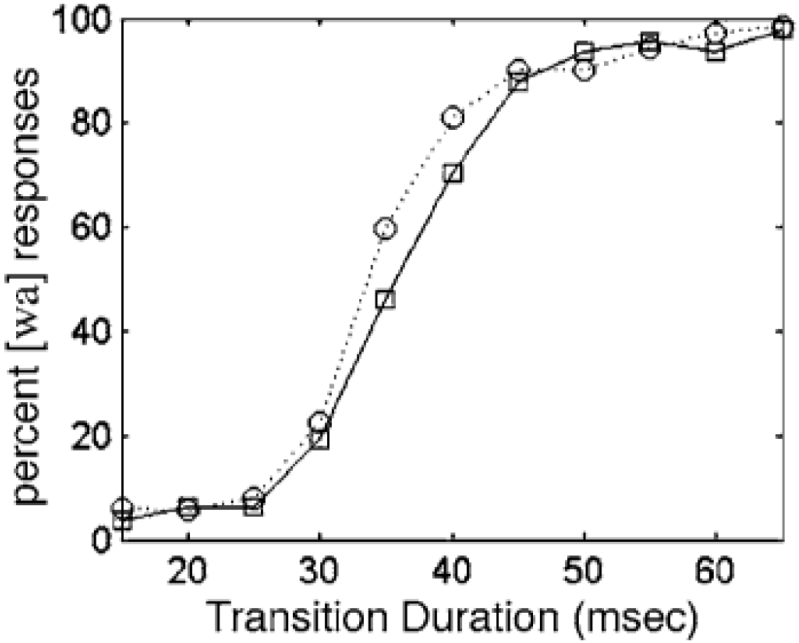


Welch et al., 2009

Rate normalization: **how?**

- Doesn't need speech!





Wade & Holt, 2005

Rate normalization: **how?**

- I’d say it needs a brain (but not aware of any ear-of-presentation tests…)

Rate normalization: **how?**

- Does it need attention?
  - Your own speech rate can change what you hear someone else say!

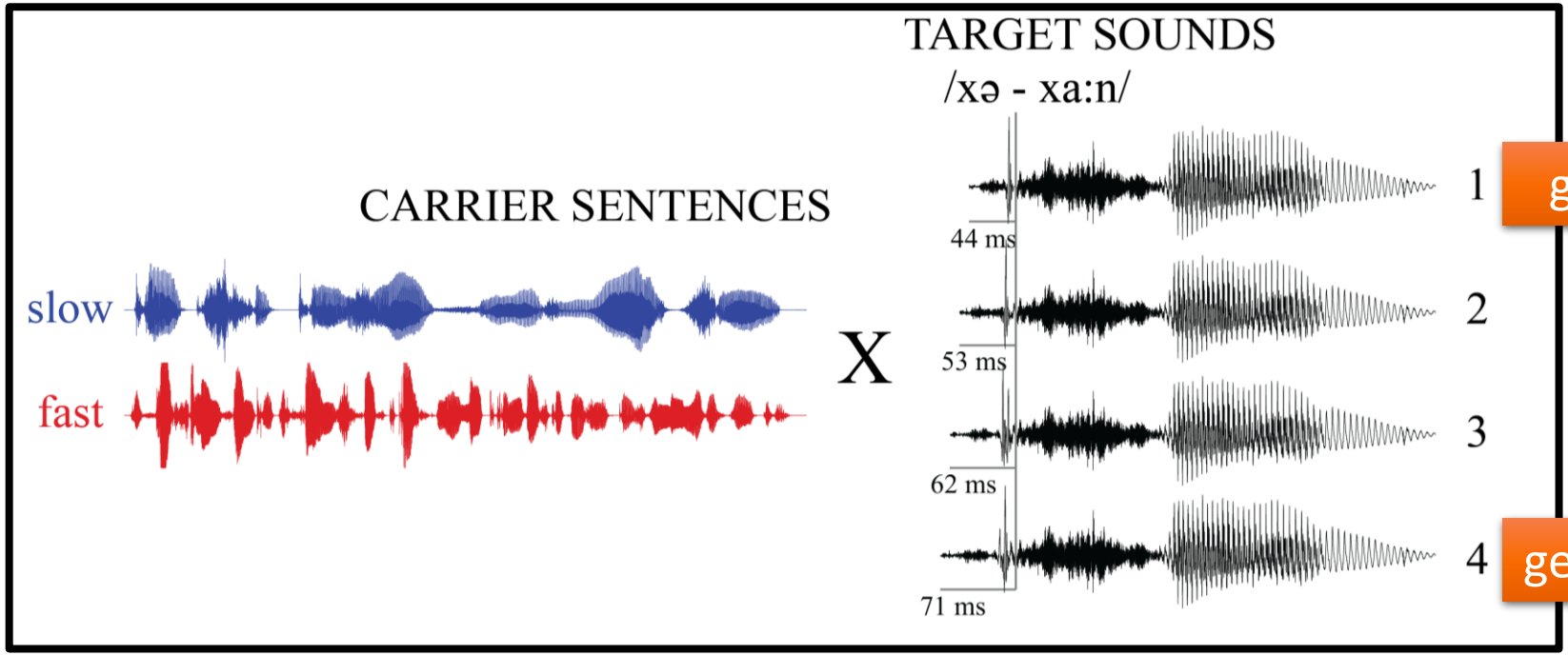
    Bosker, 2017, *JEP:LMC*
  - Reducing processing resources through dual-tasking (cognitive load) doesn't reduce rate normalization…

    …but does shrink time!

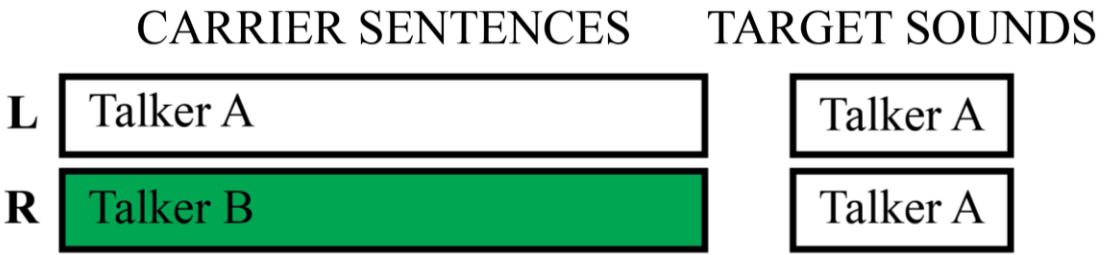
    Bosker et al., 2017, *JML*
  - Rate normalization is immune to selective attention!

    Bosker et al., 2020, *Sci Rep;* Stephens, 2022

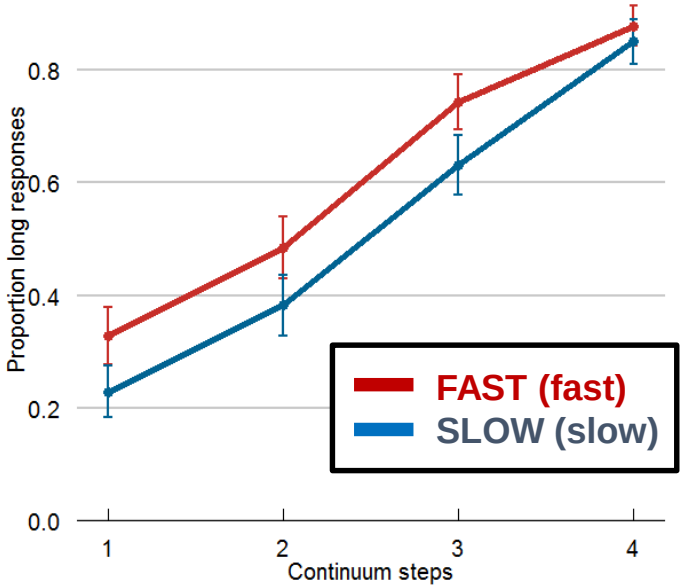
Rate normalization: **how?**



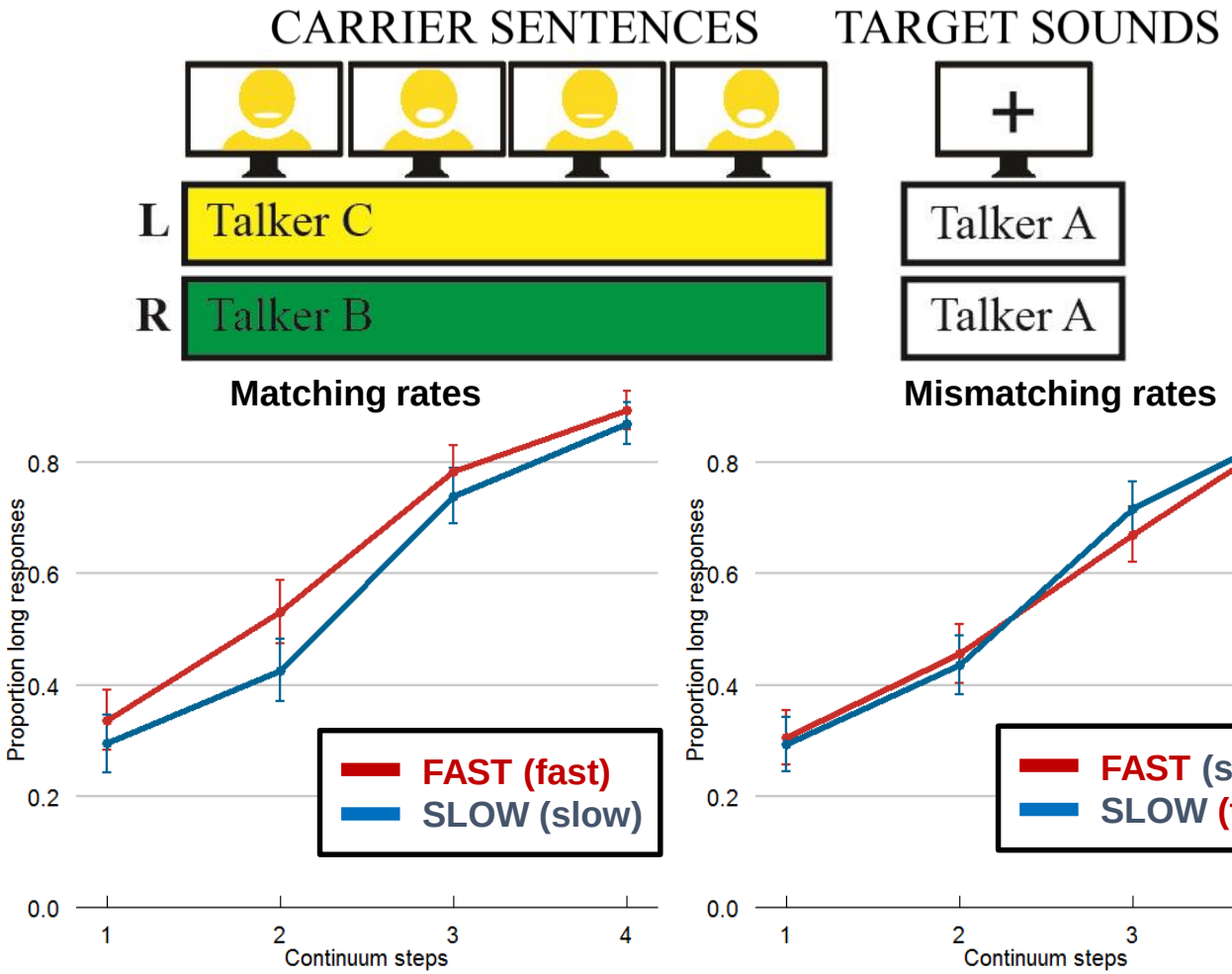
Bosker et al., 2020, *Sci Rep*



Matching rates



Bosker et al., 2020, *Sci Rep*



Bosker et al., 2020, *Sci Rep*

Rate normalization: **summary**

- We perceive target speech relative to the tempo of the surrounding acoustic context
- Rate normalization seems to involve general auditory processing levels (i.e., not speech specific), as it is observed in nonhumans, can be induced by nonspeech, and is immune to attentional influences
- That said, there are reports of more higher-level influences:
    - Foreign languages sound fast (Bosker & Reinisch, 2017)
    - Cognitive load shrinks time (Bosker et al., 2017)
    - Knowledge about a talker's habitual speech rate influences vowel perception (Reinisch et al., 2016)
    - A 'normal' speech rate sounds fast in the context of other slow speech (Maslowski et al., 2019)

Wrap-up of today

- Prosody influences segmental speech perception through acoustic context effects
- Normalization helps to overcome prosodic variability in speech
- Spectral and rate normalization involve general auditory mechanisms
- …but higher-level cognitive adjustments shape normalization perhaps at later processing stages

Next up:

- Lecture 2: *Neural tracking of 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