Speech tempo perception and deletion: Evidence from a listening experiment

Robert Lennon, Rachel Smith & Leendert Plug

LEVERHULME TRUST

UNIVERSITY OF LEEDS

University of Glasgow

Rate and Rhythm in Speech Recognition   Max Planck Institute, Nijmegen, December 2019

MOTIVATION

- Tempo is often quantified with a single rate measure, e.g.:

  Canonical syllable rate   **Surface syllable rate**

  Canonical phone rate   **Surface phone rate**
- **How do measures relate to rate as perceived by listeners?**
- This study extends work by Koreman (2006) and Reinisch (2016) on speech tempo perception, taking its main methodological cues from Koreman (2006)
- **How do listeners respond to syllable and phone deletions in estimating tempo?**

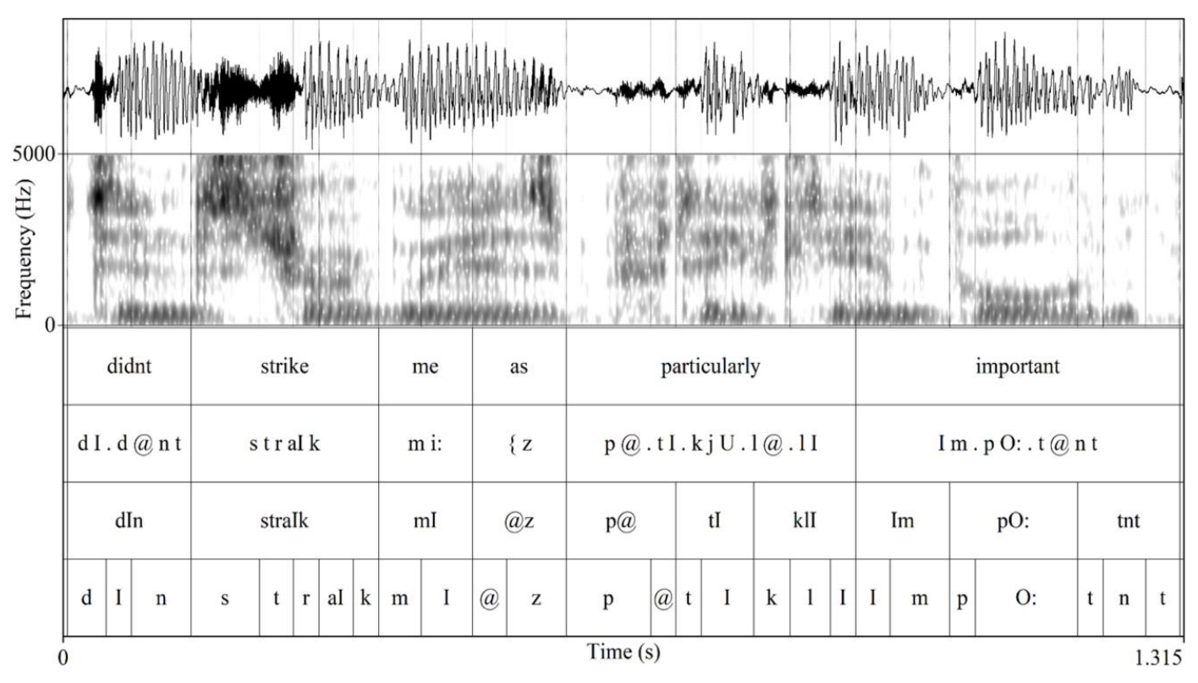
METHOD

Corpus preparation:

- 865 memory stretches (Jessen 2007) selected by Gold (2014) from DyViS corpus

  (Nolan et al. 2009; 30 SSBE males)
- Segmented in WebMAUS

  (Kisler et al. 2017)
- Syllable & phone counts extracted, canonical & surface rates calculated.



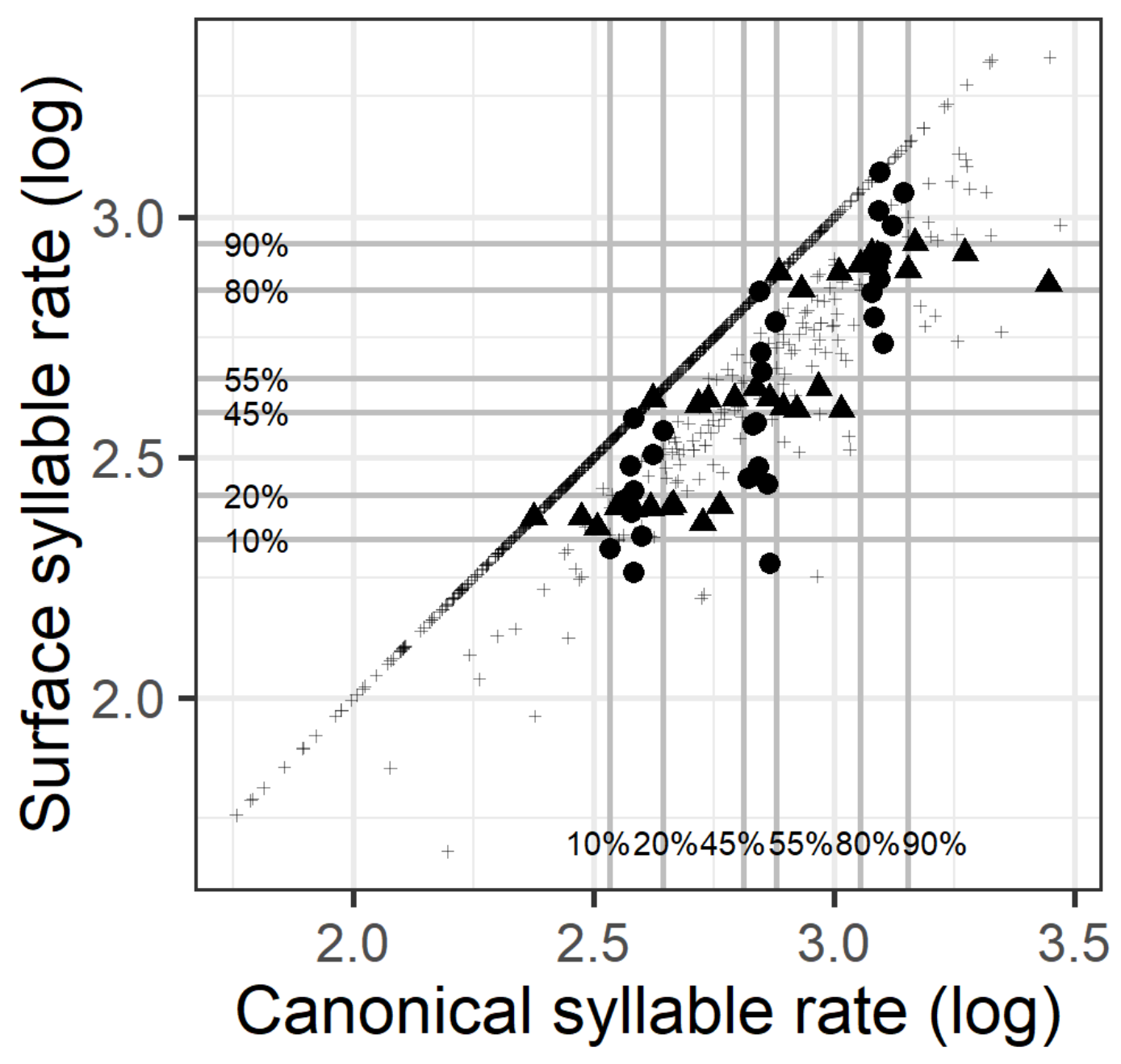
- Similar deletion rates in other corpora (English (Johnson 2004), Dutch (Van Bael et al. 2007): speakers are not unusually careful articulators
- Pairwise correlations among rates are very high: r>0.8

Stimulus selection:

- Sets of 60 stretches selected, optimized for pairwise comparisons, including:

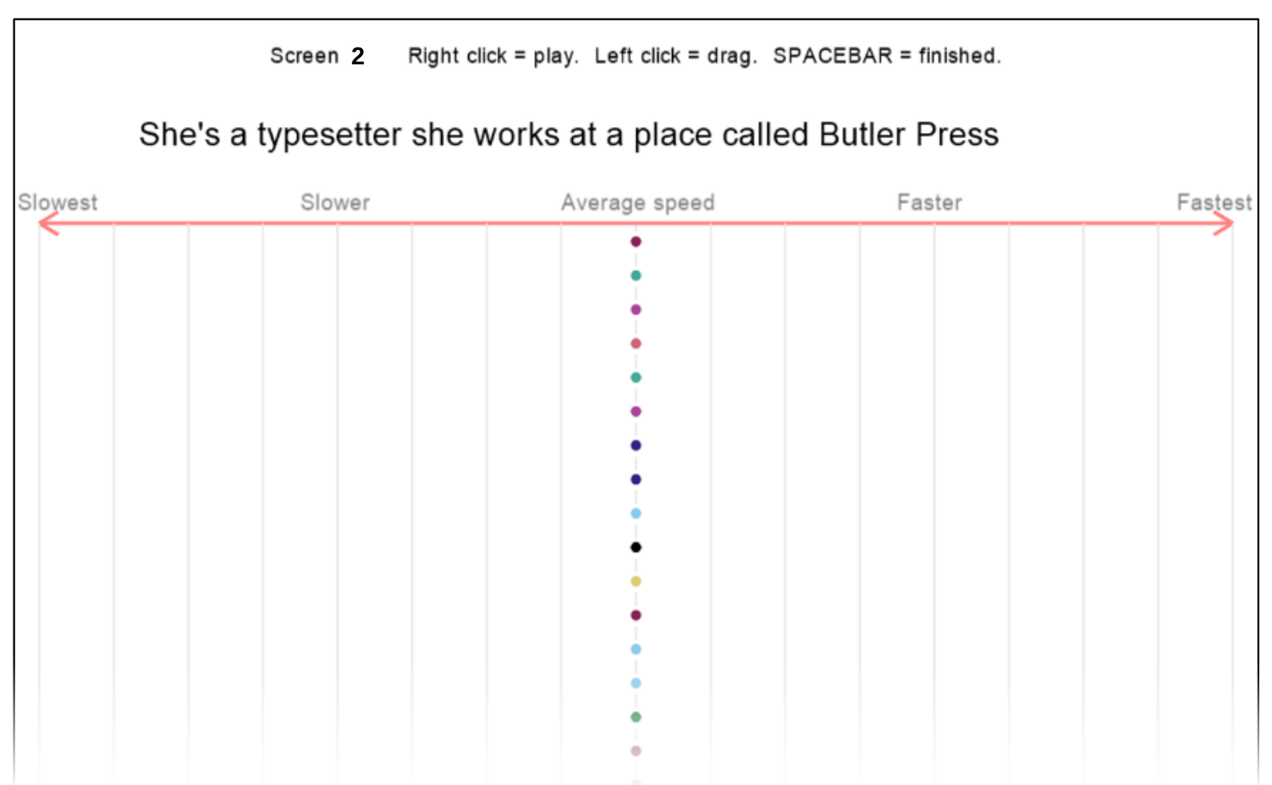
  Canonical vs surface syllable rate

  Canonical vs surface phone rate
- Each set comprises 6 subsets of 10 stimuli within which one rate is close to constant and another rate varies substantially



Experiment procedure:

- 55 listeners rated tempo in *PsychoPy2* (Peirce 2009)
- Participants clicked stimulus dots to play each stretch, then dragged to indicate perceived tempo
- 60 dots on each screen, rotated into portrait orientation


Screen 2   Right click = play. Left click = drag. SPACEBAR = finished.

She's a typesetter she works at a place called Butler Press

References

Gold 2014. *Calculating likelihood ratios for forensic speaker comparisons using phonetic and linguistic parameters*. PhD thesis, University of York.

Jessen 2007. Forensic reference data on articulation rate in German. *Science and Justice* 47, 50-67.

Johnson 2004. Massive reduction in conversational American English. *Spontaneous speech: Data and analysis. Proc. of 1st session, 10th Intl. symposium* Tokyo.

Kisler, Reichel & Schiel 2017. Multilingual processing of speech via web services. *Comp. Speech & Language* 45, 326-47.

Koreman 2006. Perceived speech rate: The effects of articulation rate and speaking style in spontaneous speech. *Journal of the Acoustical Society of America* 119, 582-596.

Nolan, McDougall, De Jong & Hudson 2009. The DyViS database: Style-controlled recordings of 100 homogeneous speakers for forensic phonetic research. *International Journal of Speech, Language and the Law* 16, 31–57.

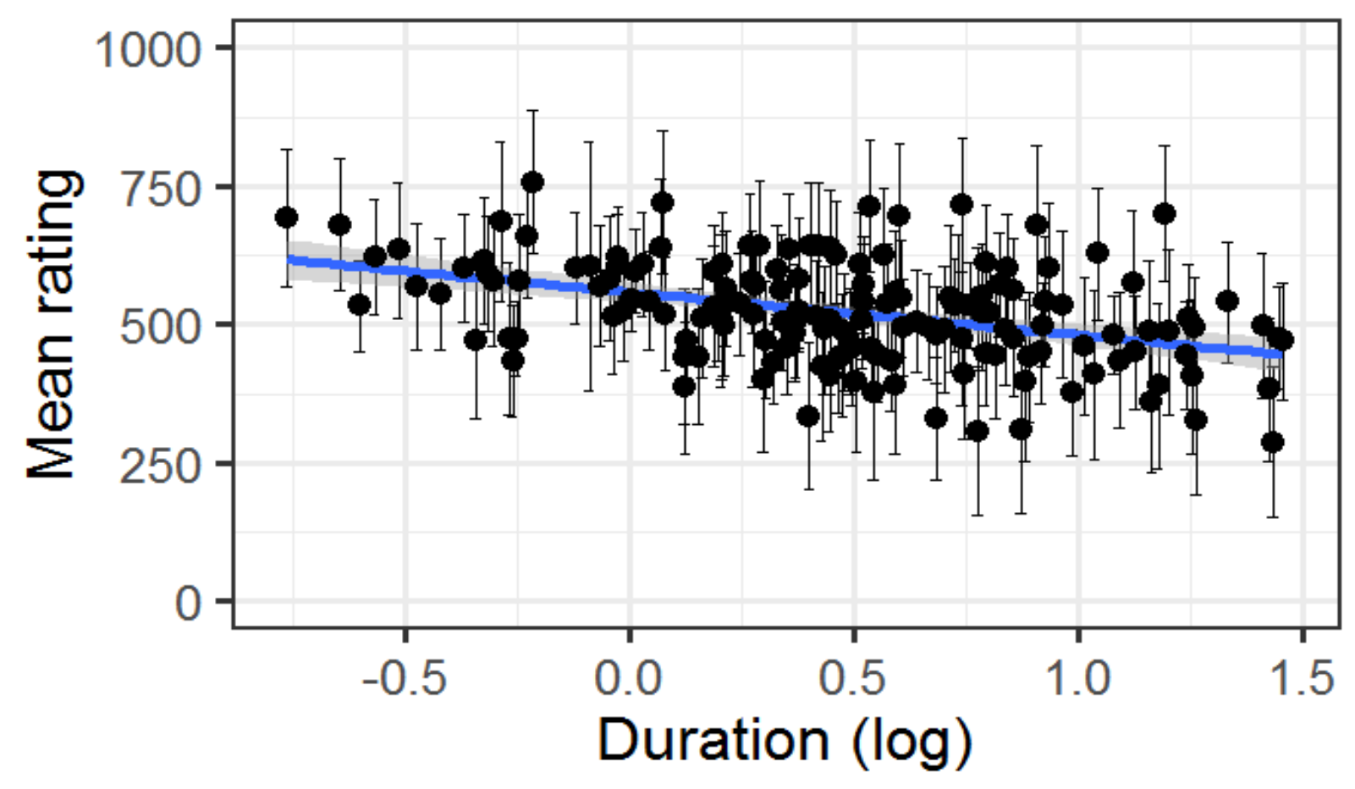
Peirce 2009. Generating stimuli for neuroscience using PsychoPy. *Frontiers in Neuroinformatics* 2.

Reinisch 2016. Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics* 78, 1203-1217.

Van Bael, Baayen, Strick 2007. Segment deletion in spontaneous speech: A corpus study using mixed effects models with crossed random effects. *Proc. Interspeech 2007* Antwerp.

RESULTS

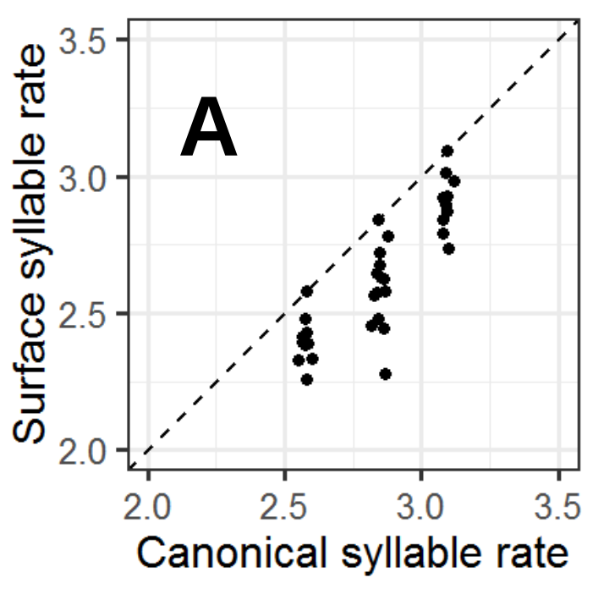
- *Mean F0* & *Mean intensity* -> higher values rated faster
- *Stimulus duration* -> negative effect (longer stimuli rated slower)



- These variables contributed to our **control model**; we then added each variable of interest, as described below

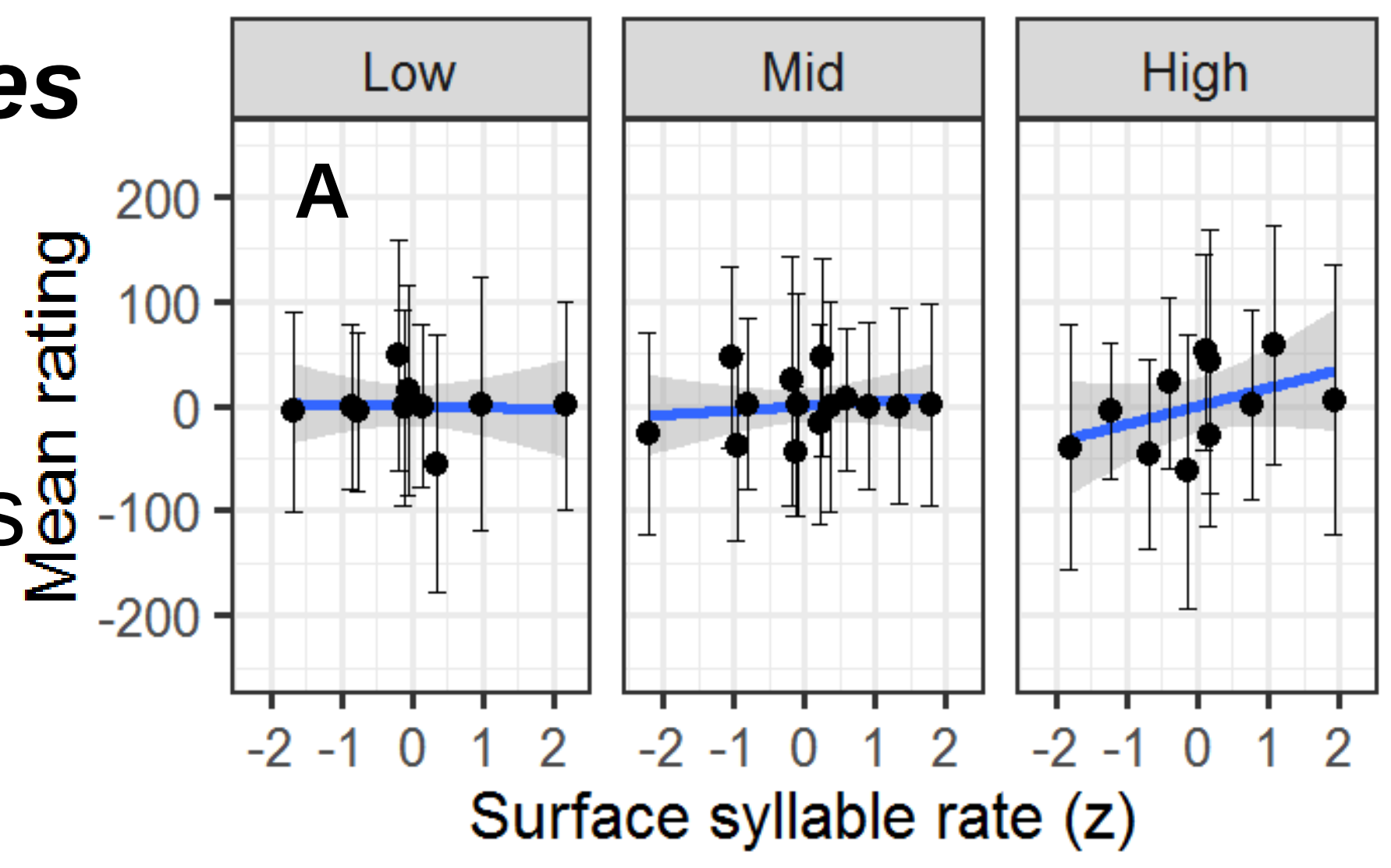
Analysis sets: One rate constant while the other varied: ‘low’, ‘mid’, ‘high’ rates

Set A: *Surface syllable rate varies*

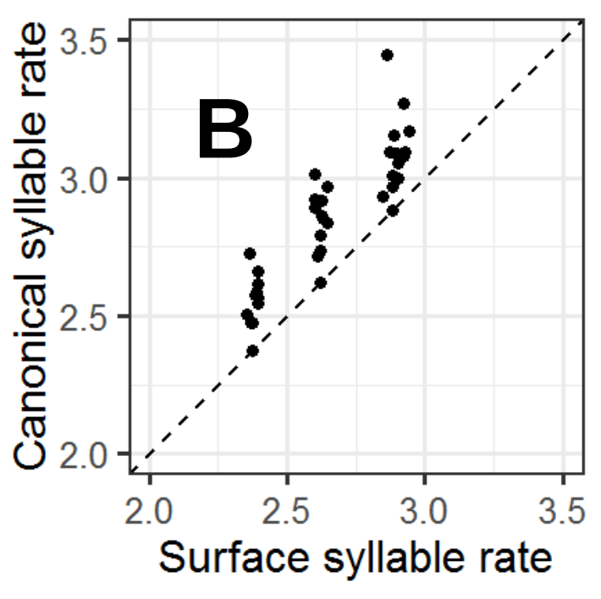


Positive effect, clearly observed at high rate.

More deleted syllables = *slower* ratings

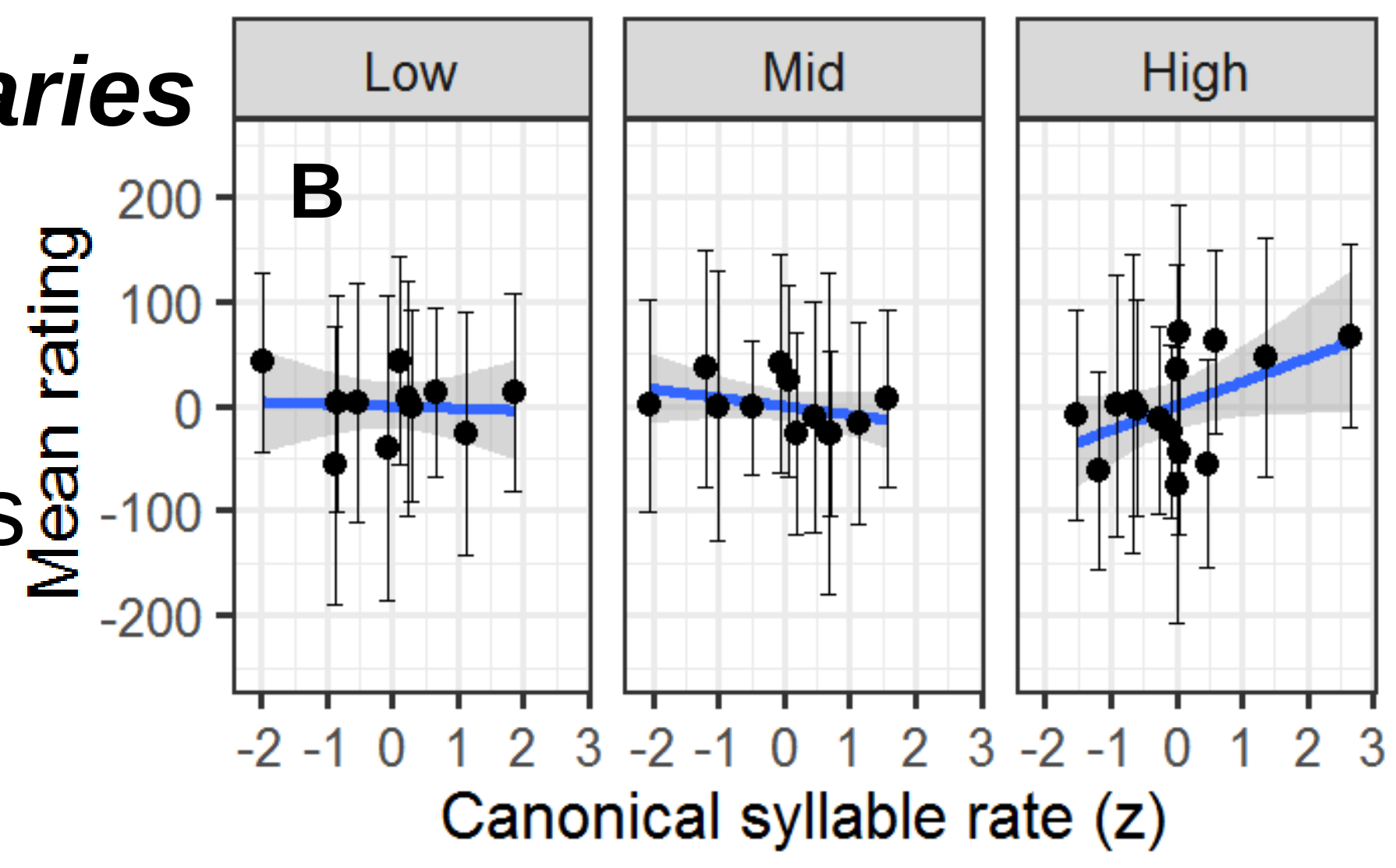


Set B: *Canonical syllable rate varies*

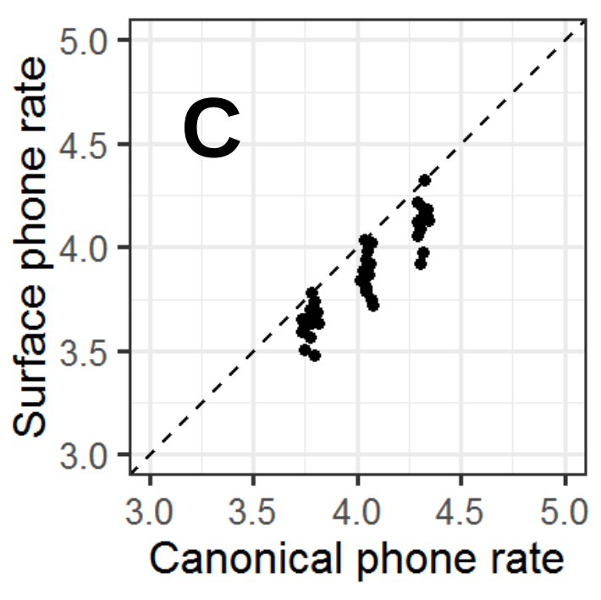


Positive effect, clearly observed at high rate.

More deleted syllables = *faster* ratings

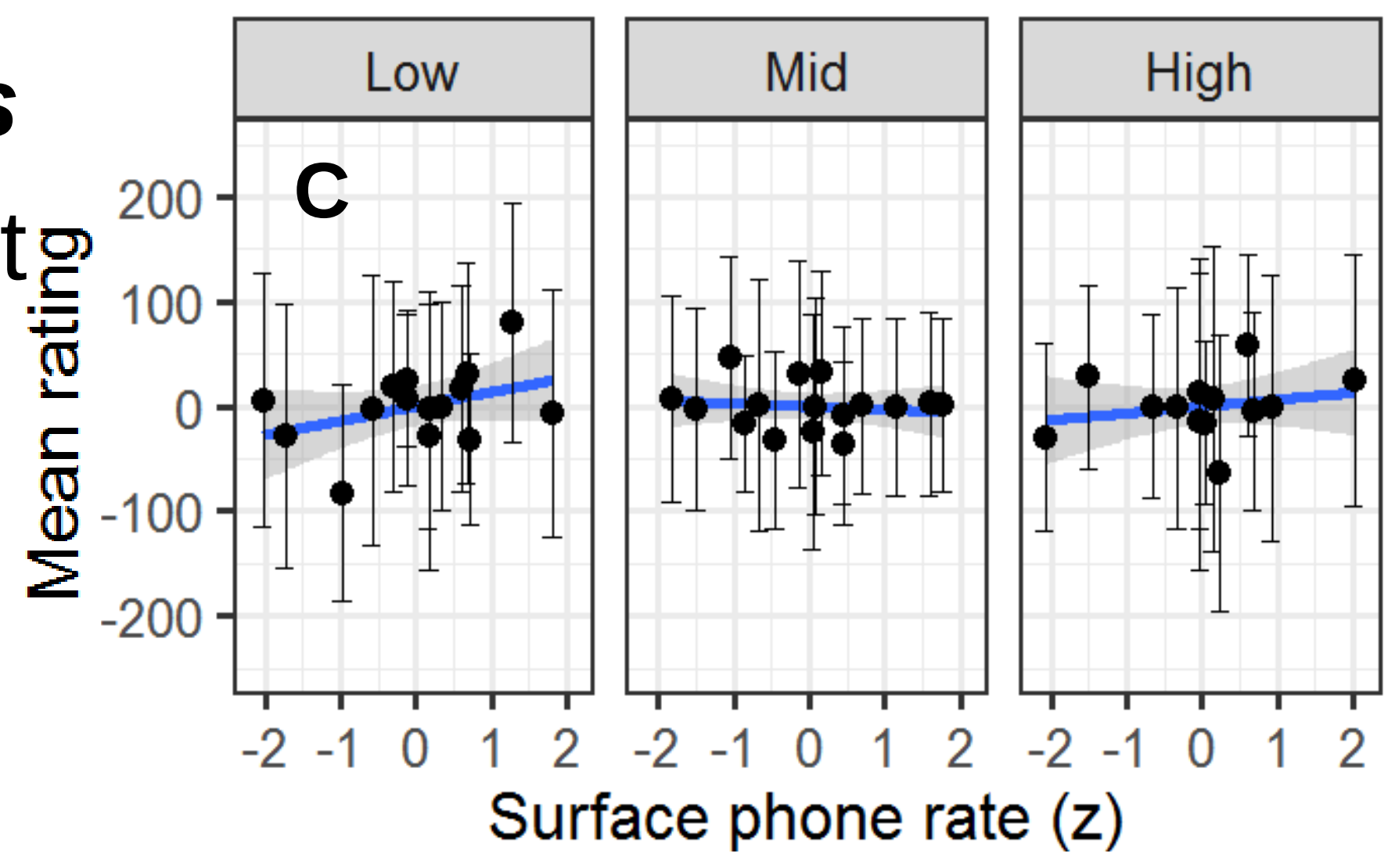


Set C: *Surface phone rate varies*

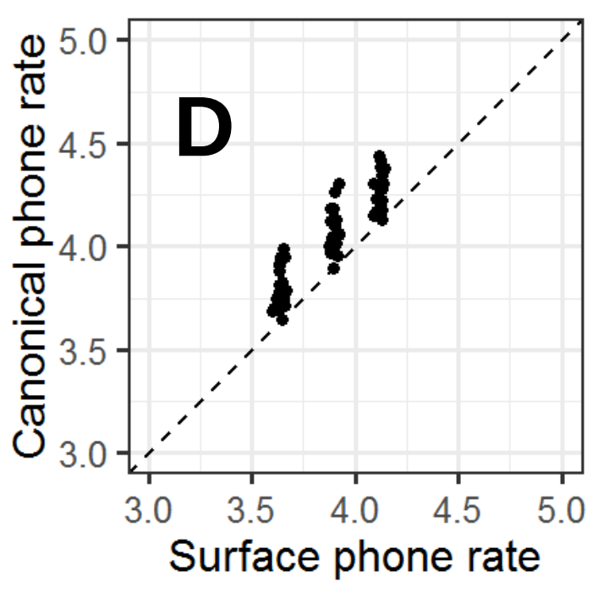


Positive effect, clear at low and high rates.

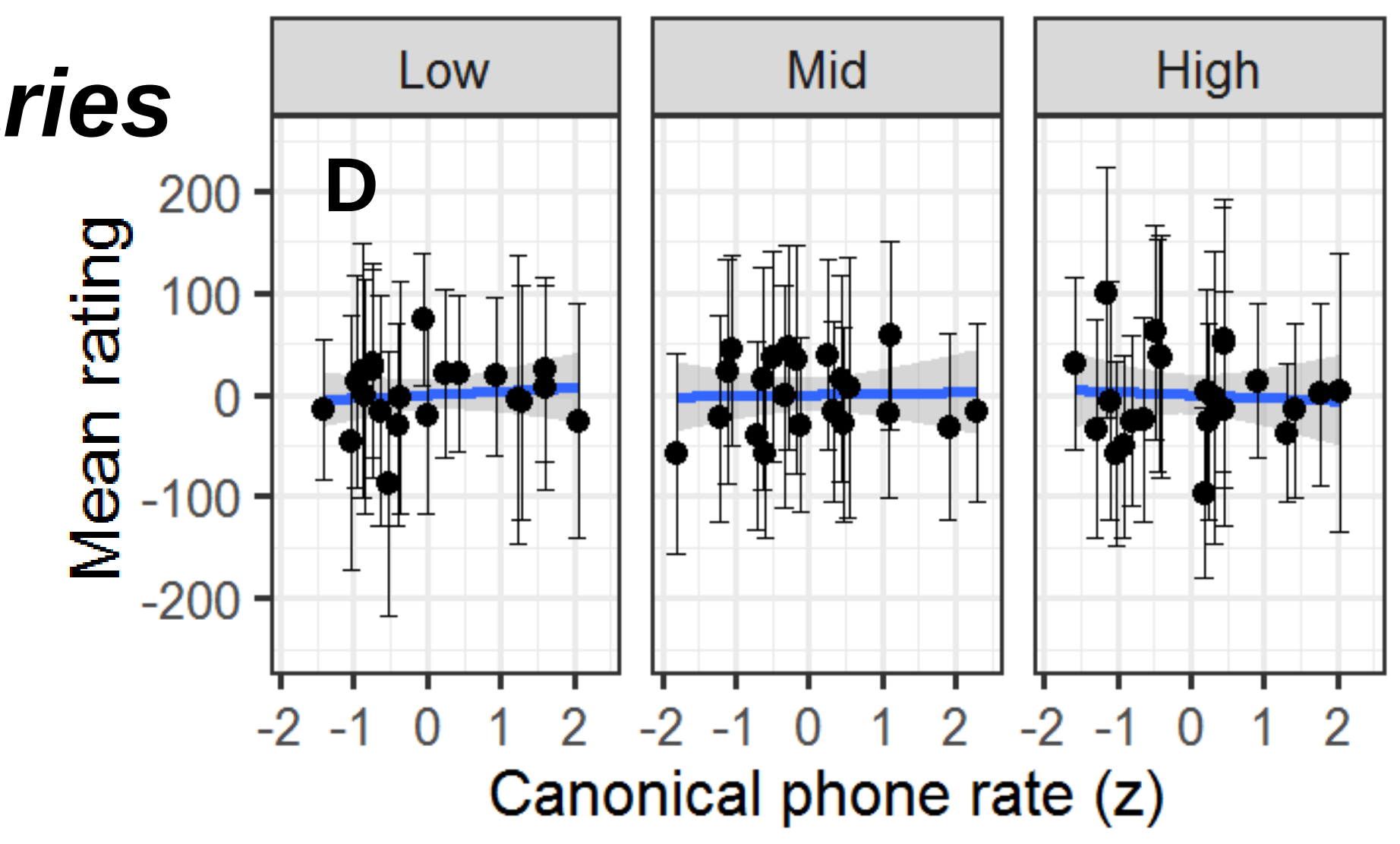
More deleted phones = *slower* ratings



Set D: *Canonical phone rate varies*



No canonical phone rate effect.



CONCLUSIONS

- Like Koreman (2006), we found that listeners do not consistently attend to some particular (measurable) temporal parameter when judging tempo.
- Listeners systematically attended to variation in both canonical and surface syllable rates – however, canonical phone rate variation was ignored.
- Phone deletions may be ignored because they can occur at all rates, whereas syllable deletions strongly indicate faster speech.
- No clear explanation for lack of sensitivity to variation in mid-tempo speech.