Does speech entrainment speed up turn-taking?

Jule Nabrotzky[1], Lars Meyer[1,2], Mathias Scharinger[3]

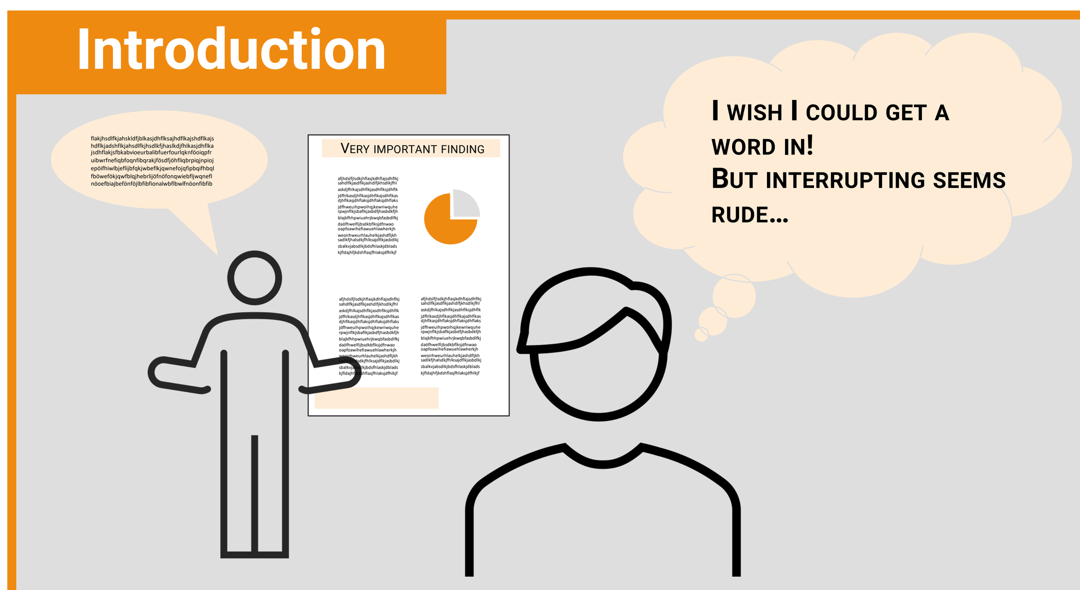
[1]Max Planck Institute for Human Cognitive and Brain Sciences, Leipzig, Germany

[2]Clinic for Phoniatrics and Pedaudiology, University Hospital Münster, Germany

[3]Philipps-University Marburg, Germany

Correspondence: nabrotzky@cbs.mpg.de

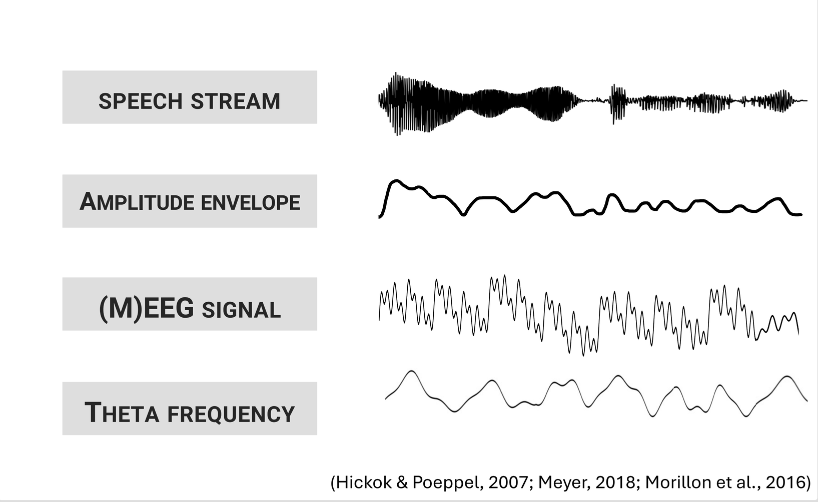
Introduction



In conversation, the silent gaps between speakers are remarkably short (~200ms).
How do listeners predict when another speaker's turn will end and then time their response correctly?

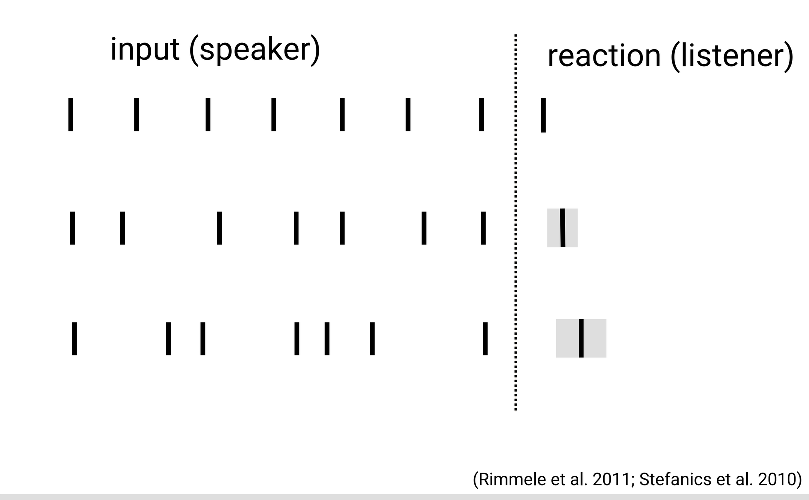
Our hypothesis:
prediction through speech tracking/speech entrainment via neural oscillations

Speech tracking: The mechanism



(Hickok & Poeppel, 2007; Meyer, 2018; Morillon et al., 2016)

Neural entrainment:



(Rimmele et al. 2011; Stefanics et al. 2010)

Research question

Is entrainment to the syllable rate contributing to the accurate timing of turn transitions?

Methods

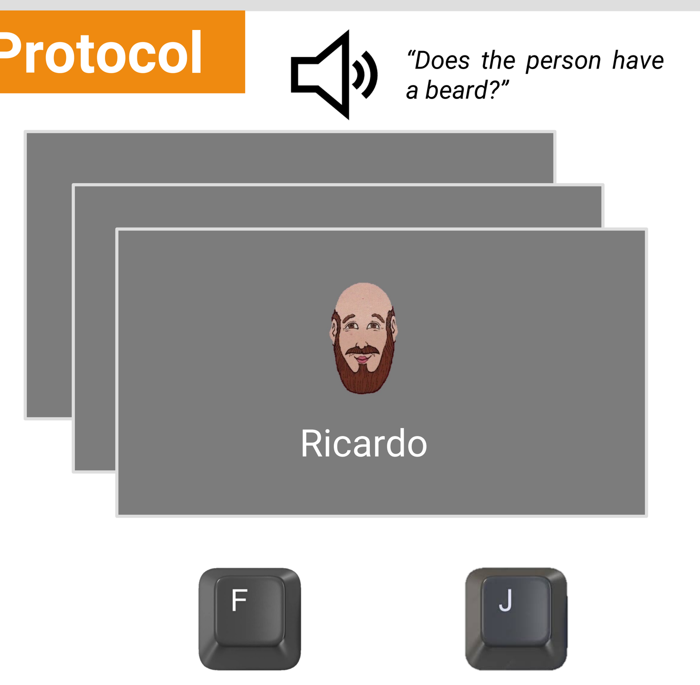
Question-answer game based on „Guess who" (behavioural)
29 participants, 23 male, aged 20-35 (M = 26.7, SD= 4.8)
with native language German and no neurological, visual or hearing impairments

270 questions in total
→ **135 x** 7±1 syllables
→ **135 x** filler (3-5 & 9-13 syllables)

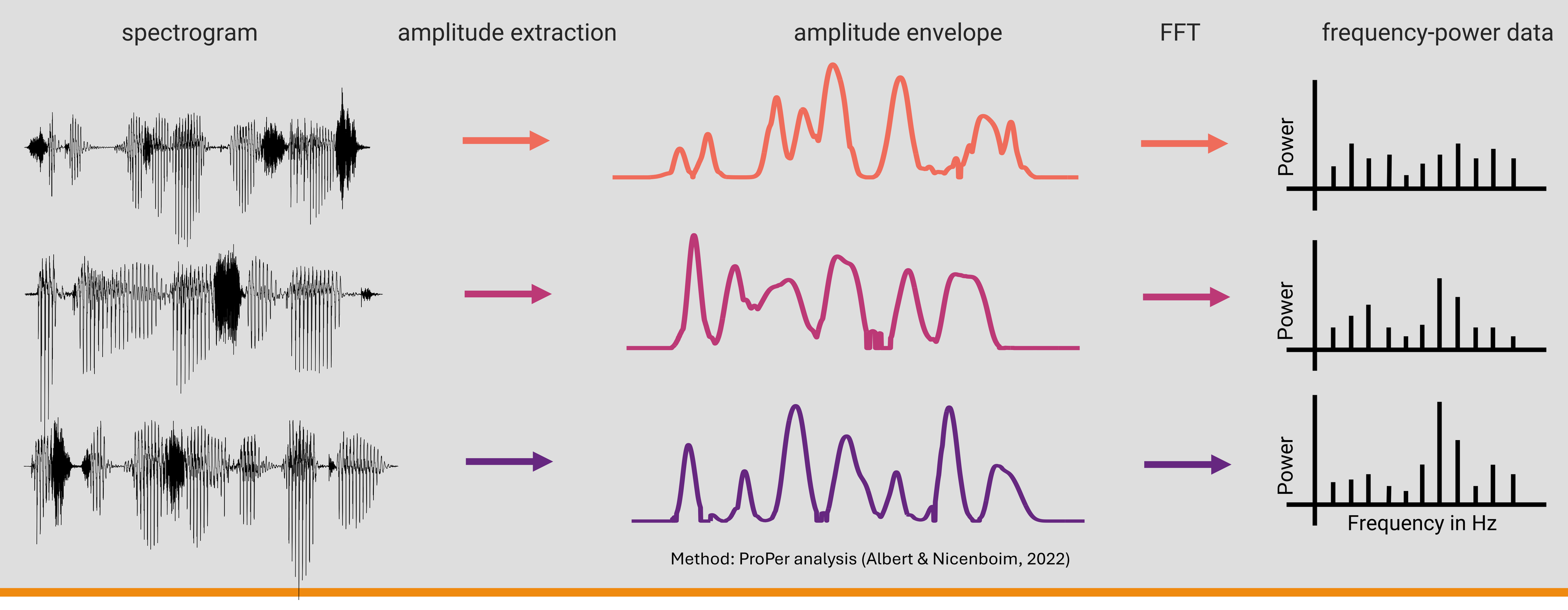
→ **135 x** correct answer: yes
→ **135 x** correct answer: no

→ presented in blocks of ~24 questions
→ **11-12** questions per person

Protocol



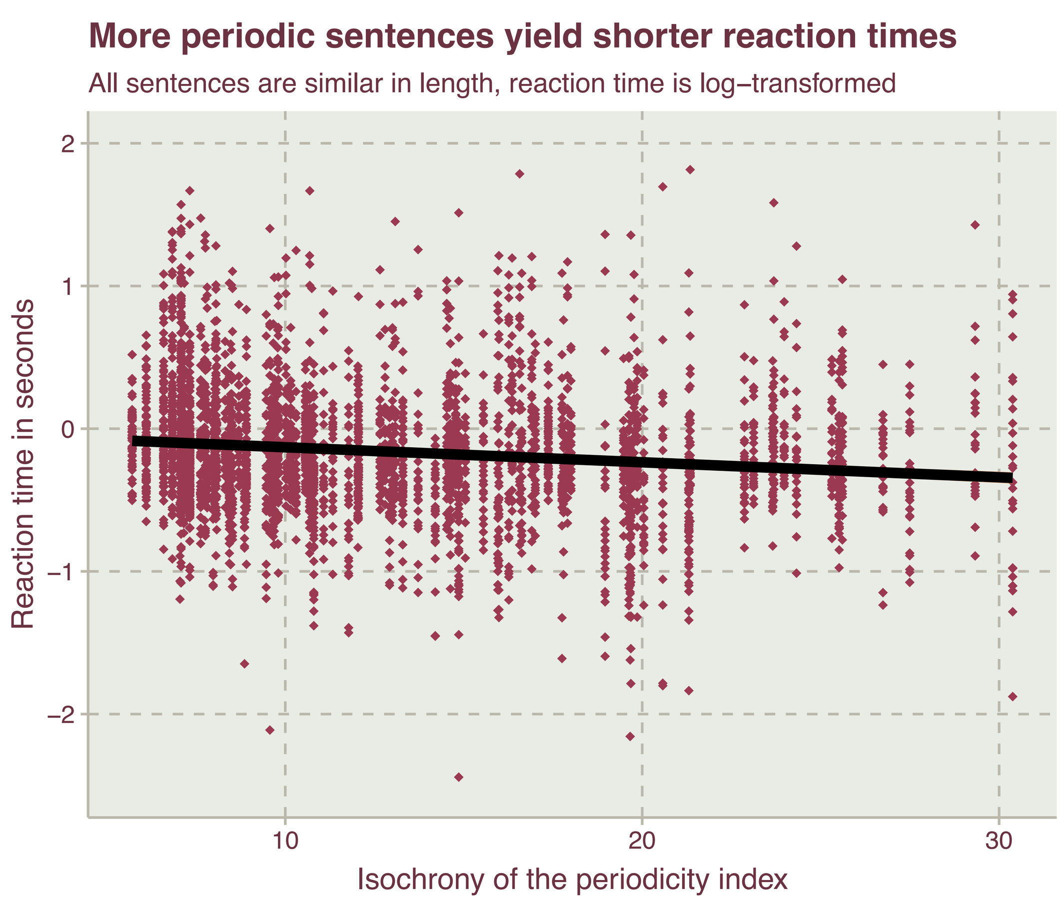
Independent variable: periodicity
measuring periodicity of amplitude envelope with time-frequency analysis



Method: ProPer analysis (Albert & Nicenboim, 2022)

Results

1

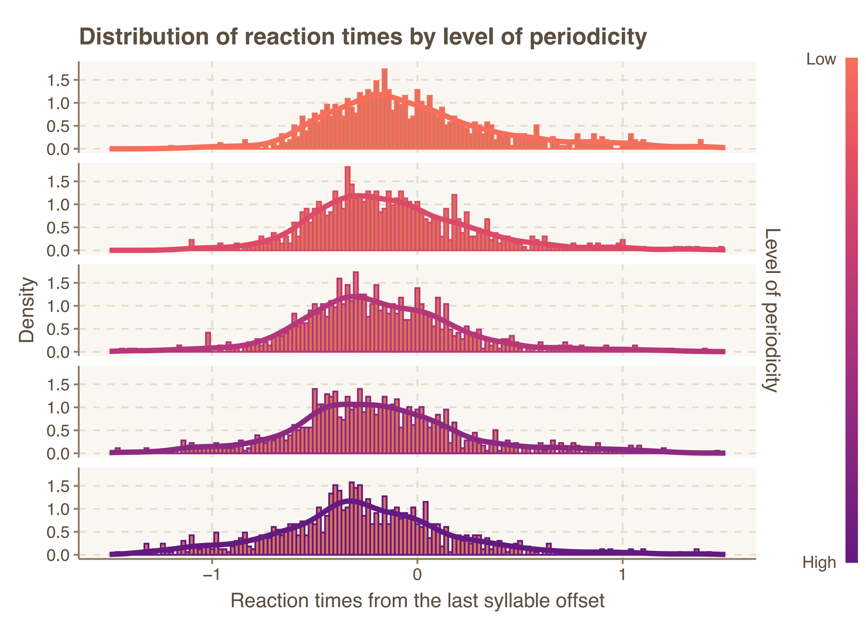


Modelling: Linear mixed model with random intercepts for item and participant, REML estimation
→ fixed effects: periodicity + sentence length * speech rate + position of relevant word
→ **significantly better fit** than reduced model without periodicity (χ^2=10.19, df = 1, p= 0.001)

Additional findings:
→ longer sentences lead to faster reaction times
→ faster speech rate leads to faster reaction times
→ earlier mention of word relevant for answering leads to faster reaction times (Brehm & Meyer, 2021; Corps, 2019; Torreira & Bögels, 2022)

☛ The more rhythmic the sentence, the faster the replies

2

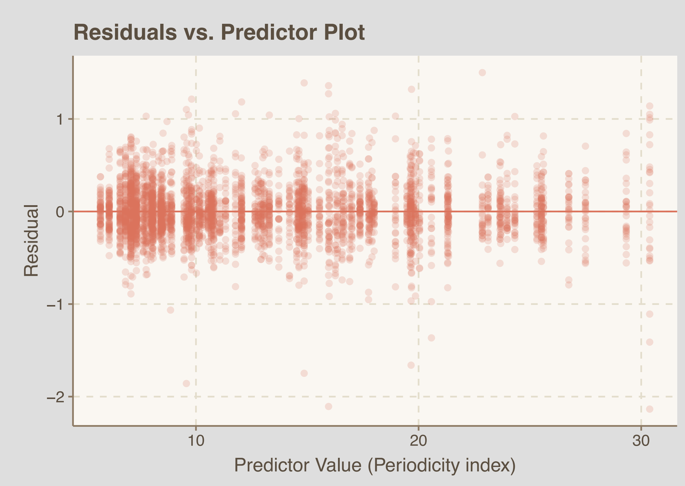


Observation: Density distribution seems to change as periodicity increases
→ Modelling squared absolute residuals over the level of periodicity confirms this
→ non-normal distributions?
→ however, bimodality not found in the data

☛ Listeners are more certain about when to reply when the sentence is rhythmic

3 **Investigating the residuals:**

HETEROGENEITY OF VARIANCE

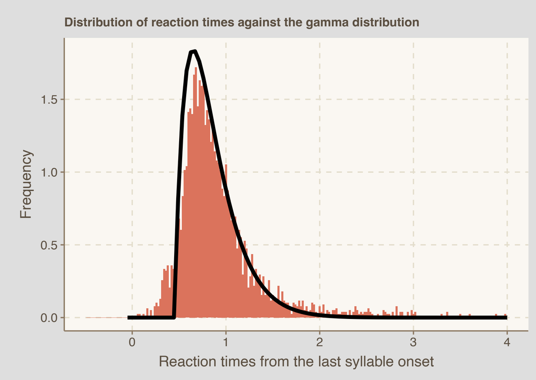


The size of residuals is negatively correlated with the predictor value (the periodicity of the sentence).

POST-HOC

Approach 1: Fitting a generalized linear mixed model on the non-transformed RT with a gamma distribution and an identity link

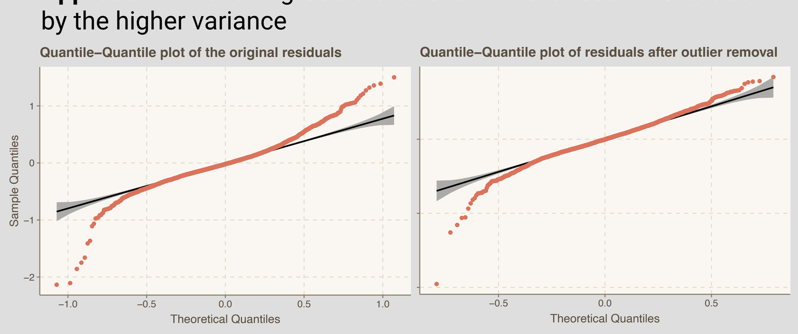
Model outcome: effect of periodicity is significant at (p=0.021)



The results are still significant, but the model does not fit well.

POST-HOC

Approach 2: Reducing outliers to check if the results is caused by the higher variance



The outcome stays significant even with outlier reduction.

Discussion

Experimental design:
- how natural is a button press?
  → button presses don't align with theta rhythm
- question-answer game vs. natural conversation
  → reduces variability due to relatively simple response planning
  → BUT: reduces taking a turn to the function of "answering"

Results and limitations:
- the sentences were not recorded with periodicity levels in mind, so the sample of the independent variable is somewhat skewed
- is a faster answer necessarily always better? induction of answer speed through participant instruction?
- longer reaction times than found in natural conversation

Future directions
- testing the entrainment to the syllable rate more directly with an MEG study
- expanding to entrainment at delta frequency (intonation phrases) and combining the two
- finding more natural, conversation-like designs that keep some control over experimental conditions
- corpus study to confirm finding

References

Albert, A., & Nicenboim, B. (2022). Modeling Sonority in Terms of Pitch Intelligibility With the Nucleus Attraction Principle. *Cognitive Science*, *46*(7), e13161.

Brehm, L., & Meyer, A. S. (2021). Planning when to say: Dissociating cue use in utterance initiation using cross-validation. *Journal of Experimental Psychology: General*, *150*(9), 1772–1799.

Corps, R. (2019). *Coordinating utterances during conversational dialogue: The role of content and timing predictions*.

Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402.

Lawrance, E. L. A., Harper, N. S., Cooke, J. E., & Schnupp, J. W. H. (2014). Temporal predictability enhances auditory detection. *The Journal of the Acoustical Society of America*, *135*(6), EL357–EL363.

Meyer, L. (2018). The neural oscillations of speech processing and language comprehension: State of the art and emerging mechanisms. *European Journal of Neuroscience*, *48*(7), 2609–2621. https://doi.org/10/gddhg4

Morillon, B., Schroeder, C. E., Wyart, V., & Arnal, L. H. (2016). Temporal Prediction in lieu of Periodic Stimulation. *The Journal of Neuroscience*, *36*(8), 2342–2347.

Ringer, H., Schröger, E., & Grimm, S. (2023). Neural signatures of automatic repetition detection in temporally regular and jittered acoustic sequences. *PLOS ONE*, *18*(11), e0284836.

Torreira, F., & Bögels, S. (2022). Vocal reaction times to speech offsets: Implications for processing models of conversational turn-taking. *Journal of Phonetics*, *94*, 101175.

1st IMPRS CoNI Summer School, 2024