Tracking speaker-specific speech rate: Habitual vs. local influences on English stop voicing

Connie Ting[1,2] & Yoonjung Kang[3]

[1]McGill University, [2]Centre for Research on Brain, Language and Music, [3]University of Toronto

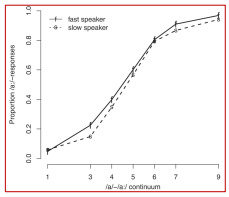
1. Background

- Speech rate is a source of variation that creates differences in the realization of durational acoustic cues
  - In **fast speech**, overall duration is **shorter**
  - In **slow speech**, overall duration is **longer**
- Therefore, an ambiguous phonetic segment is long relative to surrounding fast speech, but short relative to surrounding slow speech

- *Local* speech rate effects in perception have been observed for:

Temporal Cue	Fast speech	Slow speech
Vowel Duration	More /a:/	Less /a:/
Stop VOT	More /p/	Less /p/

- Reinisch (2016) tested *speaker-specific* effects of speech rate on listeners' vowel length perception in German
  - Listeners heard a 2-minute dialogue between two female German speakers, varying in:
    - Rate:
      - slow = 10% slower than mean
      - fast = 15% faster than mean
    - Order (first vs. second)
  - Listeners categorized words of minimal Pair continua differing in /a/-/a:/ contrast



- Results showed that listeners are able to track speaker-specific rate information to facilitate vowel length perception in German
  - A study aimed to replicate Reinisch(2016)'s findings with English listeners' perception of a consonantal contrast (Ting & Kang, 2023)
  - Results showed no speech rate effect for consonant stop voicing when the target segment was not present in the dialogue

Gap in previous work:

❖Reinisch (2016) did not test whether the speech rate effect exists when listeners are not exposed to target stimuli in the dialogue

❖Ting & Kang (2023) did not test the effect in more local contexts and included only a single word pair, which could contribute to a less natural task

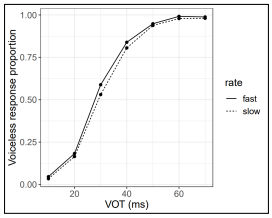
4. Results

Analysis

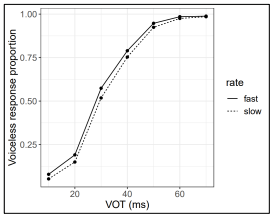
- A logistic mixed-effects model was run for each experimental condition (words in isolation, sentential context):
  - Dependent variable: voiceless vs. voiced responses
  - Fixed effects:
    - VOT
    - Speech Rate
  - Random effects:
    - Subject
    - Speaker

Results

ID TASK 1 – WORDS IN ISOLATION



ID TASK 2 – WORDS IN CARRIER SENTENCE



✦ Mixed-effects logistic regression model revealed a significant effect of speech rate for both tasks

✦ Results of ID Task 1 suggest listeners did make use of speaker-specific speech rate information in VOT perception

✦ Results of ID Task 2 suggest speech rate effect is also observed in a local speech rate context

2. Current Study

The current study:

❖ Aims to replicate previous studies using VOT (/bi/-/pi/ continuum) with modifications:

1. increasing the rate difference
2. using multiple word pairs
   big-pig bye-pie
   doe-toe dime-time
3. including two identification tasks:
   - ➢ Isolated stop-initial syllables manipulated along a VOT continuum
   - ➢ The same syllables embedded in a fast or slow carrier sentence

3. Methods

Participants

- 80 self-reported native American English speakers recruited through Amazon Mechanical Turk

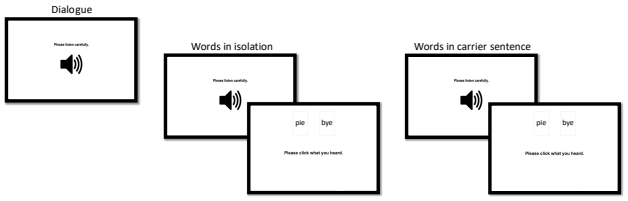
Stimuli

- 2 male speakers recorded both roles of a 460-word dialogue between 2 speakers
- Rate manipulation (relative to average within pairs):
  - Fast = 20% shorter
  - Slow = 20% longer
- Identification task VOT manipulation:
  - /b/-/p/, /t/-/d/ continua
  - 10-70 ms in 7 equal steps
  - Constant vowel duration = average within pairs

- **Task 1:** 112 words in isolation (= 7 VOT steps * 4 word pairs * 2 speakers * 2 reps)
- **Task 2:** 112 sentences (= 7 VOT * 4 word pairs * 2 speakers * 2 sentence rates)

Procedure

- Each listener heard one of four versions of the dialogue (each version was heard by 20 participants)
- After the dialogue, listeners categorized:
  - Isolated /bye-pie/ syllables spoken by both speakers within the pair and indicated what they heard by button click
  - The same syllables embedded in a fast or slow carrier sentence



5. Conclusions

- **speaker-specific speech rate** in this experiment affected listeners' perception of the **English voicing contrast**
- rate normalization in listeners' perception of consonantal contrasts **can occur in both local (sentential) and habitual (dialogue) contexts**
- The presence of the rate effect with a more exaggerated rate difference (20% in this study compared to 15% in the previous replication with English stop voicing) indicates a **more exaggerated rate difference is required to observe a speech rate effect given a dialogue context**
- Taken together with the results of Reinisch (2016), these findings suggests **speech rate normalization differs across different durational contrasts**
- It remains unclear whether speaker-specific rate effects can be found for a vowel length contrast using a dialogue which omits all instances of the target stimuli

Acknowledgements

This research was supported by the Social Sciences and Humanities Research Council of Canada (#435-2020- 0209) and the Centre for Research on Brain, Language and Music.