The Effect of Habitual Speech Rate on Speaker-specific Processing in English Stop Voicing Perception

Connie Ting & Yoonjung Kang
University of Toronto

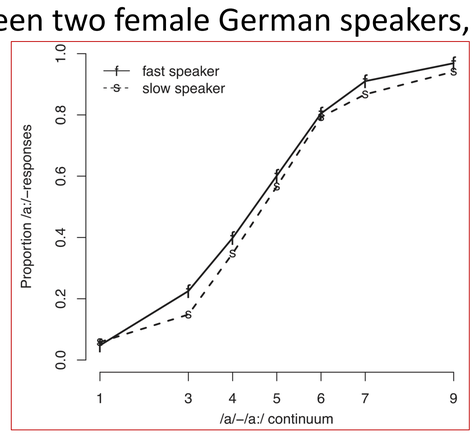
1. Background

- 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 (fast vs. slow)
    - 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.

Gap in previous work:

❖ Results of speaker-specific rate effects are restricted to the context of vowel duration contrasts.

The current study:

❖ Aims to replicate the study by Reinisch (2016) using VOT (/bi/-/pi/ continuum).

Hypothesis:

❖ **Speech Rate Effect:** Fast speech rate condition will elicit more /pi/ responses compared to slow speech rate condition.

2. Methods

EXPERIMENT 1 – SPEECH RATE EFFECT

Participants

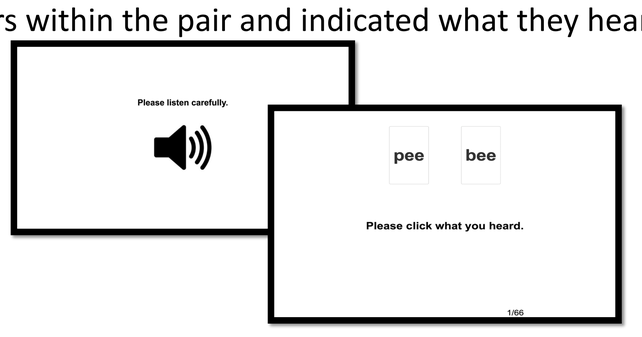
- 208 self-reported native American English speakers recruited through Amazon Mturk

Stimuli

- 4 speakers (2 male) recorded both roles of a 2-minute dialogue between 2 speakers
- Rate manipulation (relative to average within pairs):
  - Fast = 15% slower
  - Slow = 10% longer
- Identification task VOT manipulation:
  - /bi/-/pi/ continuum
  - 0-50 ms in 11 equal steps
  - Constant vowel duration = average within pairs
  - 66 tokens = 11 VOT steps * 2 speakers * 3 repetitions

Procedure

- Each listener heard one of eight versions where speakers were matched in gender (2 roles * 2 rates * 2 genders)
- After the dialogue, listeners heard isolated /bi-pi/ syllables spoken by both speakers within the pair and indicated what they heard by button click.



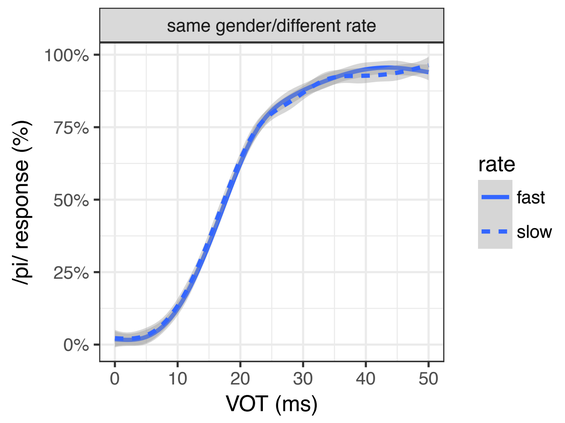
Analysis

- Participants were excluded if they did not show a significant effect of VOT, or showed a significant effect of VOT in the opposite from expected direction (*less* /pi/ responses for longer VOT) (n=49).
- A logistic mixed-effects model was run for each experimental condition:
  - Dependent variable: /pi/ vs. /bi/ responses
  - Fixed effects:
    - VOT
    - Speech Rate
  - Random effects:
    - Subject
    - Speaker

3. Results and Follow-up Studies

EXPERIMENT 1 – SPEECH RATE EFFECT

- ✦ Mixed-effects logistic regression model revealed no significant effect of speech rate.
- ✦ Results suggest listeners did not make use of speaker-specific speech rate information in VOT perception.



EXPERIMENT 2 – GENDER EFFECT

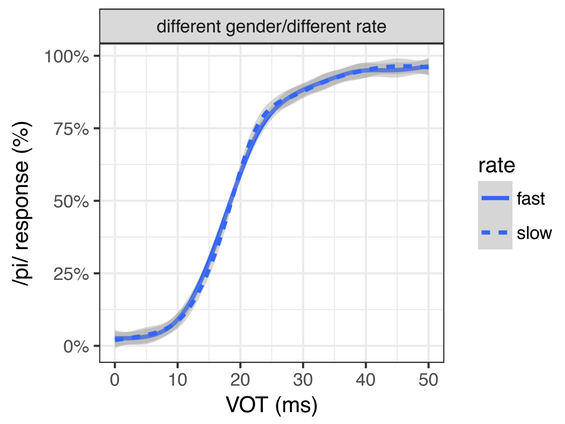
- Lack of speech rate effect in Experiment 1 may be a result of listeners having difficulty distinguishing between speakers matched in gender and approximate age.
- Perceived gender differences may also influence changes in perception due to speech rate.

Participants: 191 self-reported native American English speakers who did not participate in Experiment 1.

Stimuli, procedure, and analysis were the same as Experiment 1 except that speakers within each pair were mis-matched in gender.

Results

- ✦ No significant rate effects were found even when speakers differed both in gender and rate.
- ✦ Lack of speaker-specific rate effect is not likely due to failure to distinguish two speakers in conversation.



EXPERIMENT 3 – INTRASPEAKER VARIATION

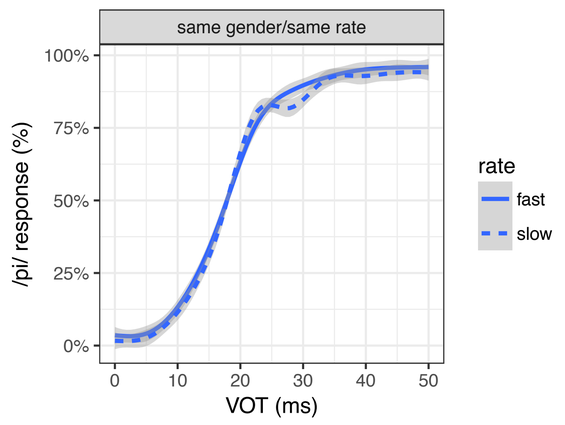
- Lack of speech rate effect in Experiments 1 and 2 may be a result of listeners still being sensitive to speech rate in the dialogue they are exposed to but aggregating speech rate information from the two speakers in the dialogue.
- If there is an effect of exposure to conversation, we expect to find differences between two dialogue conditions:
  1) both speakers with fast speech rate
  2) both speakers with slow speech rate

Participants: 209 self-reported native American English speakers who did not participate in Experiments 1 and 2.

Stimuli, procedure, and analysis were the same as Experiment 1 except that speakers within each pair were matched in speech rate.

Results

- ✦ No significant results were found across conditions when both speakers of the dialogue spoke at the same rate.
- ✦ Speech rate of the dialogue did not affect VOT perception.



4. Conclusions

- Results from the current study showed listeners did not make use of speaker-specific speech rate information in English VOT perception, inconsistent with findings regarding speaker-specific speech rate effects in vowel perception (Reinisch, 2016).
- This suggests a difference between perception of vowels and consonants which may be due to differences in speech rate induced change in vowels vs. consonants.

- Current results are in line with English production results showing rate-independent VOT categories in spontaneous speech production (Nakai & Scobbie, 2016).
- Additionally, studies have found that the VOT-speech rate correlation across speakers is not always consistent (Benjamin, 1982; Allen et al., 2003).
- If speech rate in a habitual context is not a reliable cue to VOT, any adjustment in perception due to speech rate may not be necessary.

- This can be tested with other durational contrasts that may have a greater degree of overlap due to rate variation.
- Effects of intra-speaker and gender variation on speaker-specific speech rate tracking can be further explored in vowel perception.

Acknowledgements

Thanks to Na-Young Ryu and Hyoung Seok Kwon for assistance in running online experiments.
Funding: University of Toronto Scarborough Research Competitiveness Fund.