2021

# The Role of Talker in Adjusting for Different Speaking Rates in Speech Perception

Chloe M Sharpe
chloe.sharpe@louisville.edu

Christian Stilp

# The role of talker in adjusting for different speaking rates in speech perception

Chloe M. Sharpe and Christian E. Stilp
Department of Psychological and Brain Sciences
University of Louisville

## INTRODUCTION

- **Context** is anything that can be perceived outside of what you're focusing on.
- **Context Effects** are what happens when context modifies the experience of your focus. They are very prevalent in perception.

- **Types of acoustic context effects**
  - **Spectral Contrast Effects (SCEs)** – induced by the frequencies of sounds
    - The difference between the frequency content of two sounds is perceived to be larger than they are.
    - If the frequency of the context is high, the target sound will be perceived as lower frequency and vice versa. (Ladefoged and Broadbent, 1957)

  - **Temporal Contrast Effects (TCEs)** – induced by speech rate
    - Context spoken at a fast rate will make the target sound slower.
    - Context spoken at a slow rate will make the target sound faster. (Summerfield, 1981)

| Context | More likely to perceive |
|---|---|
| Sentence (unmodified) | "deer" or "tier" |
| Sentence spoken at a fast rate | "tier" (long VOT) |
| Sentence spoken at a slow rate | "deer" (short VOT) |

- **Talker Variability** is the concept that there is great acoustic variability amongst different talkers.
  - When the talker is consistent from trial to trial, speech perception is faster and/or more accurate.
  - When there is a new talker in each trial, performance is slower and/or less accurate. (Stilp & Theodore, 2020)

- **Talker variability's effect on SCEs**
  - Talker variability has been shown to diminish SCE magnitudes.
  - Listeners exhibit smaller SCEs following multiple talkers and larger SCEs following a single talker. (Assgari & Stilp, 2015)

- However, it is unknown whether Talker Variability impacts TCEs.

- **Prediction**
  - TCE magnitudes will decrease following context sentences spoken by 200 different talkers compared to context sentences spoken by a single talker.

## PARTICIPANTS

- 41 undergraduate students in the Department of Psychological and Brain Sciences signed up through SONA and received course credit for their participation
  - All reported being native English speakers with no known hearing impairments

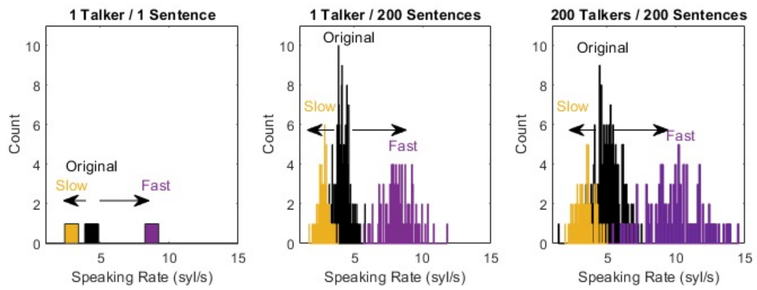
- The experiment was conducted online using Gorilla (Anwyl-Irvine et al., 2020)

## METHODS

**Stimuli**

Context Sentences

- *One Talker / One Sentence*: One adult man saying, "This time, I want you to click on the word" (duration = 2293ms)
- *One Talker / 200 Sentences*: One adult man reading 200 sentences (mean duration = 1793ms)
- *200 Talkers / 200 Sentences*: 138 men and 62 women reading their own individual sentence (mean duration = 2248ms)
- Each sentence's speaking rate was manipulated in Praat using the PSOLA algorithm. (Boersma & Weenink, 2020)
  - In each condition testing 200 sentences, half were adjusted to speak faster, and half were adjusted to speak slower.
  - In the *One Talker / One Sentence* condition, slow and fast renditions were created




Target Consonants

- Series of 10 stimuli morphed from "deer" to "tier"
- Generated in Praat by varying the voice onset time (VOT; duration the vocal cords are not vibrating during production)
- The voiceless interval at the beginning of "deer" was made increasingly longer until a clear "tier" was heard. (Winn, 2020)

**Procedure**

- *Informed Consent*: participants read and signed a consent form
- *Headphone Screen*: participants discerned which of 3 tones were the quietest, tones sounded different over headphones vs. speaker (n=9 removed from analyses)
- *Practice:* 20 sentences paired with endpoint consonants; >80% categorization accuracy needed to pass (no one removed from analyses)
- *Test:* Each trial presented a context sentence, then a target word, which listeners labeled as "deer" or "tier". Three blocks were tested (One Talker / One Sentence, One Talker / 200 Sentences, 200 Talkers / 200 Sentences) each with 200 trials (n=9 removed from analyses due to not finishing). Participants had to maintain 80% accuracy on endpoint stimuli (n=3 removed from analyses).

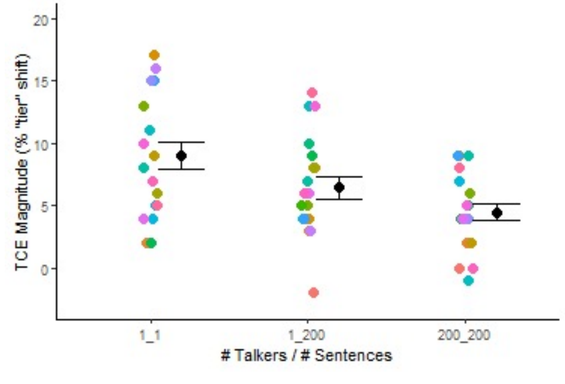
## RESULTS

- TCEs were calculated as the percent "tier" responses following fast sentences minus the percent "tier" responses following slow sentences.
  - Larger TCEs in *One Talker / One Sentence* condition and smaller TCEs in *200 Talkers / 200 Sentences* condition

- RMANOVA was statistically significant ($F(2,38)=6.05$, $p<.05$)

- Three post-hoc Bonferroni corrected paired sample *t*-tests
- *One Talker / One Sentence* and *One Talker / 200 Sentences*: not significant ($t(19)=1.59$, $p=0.128$)
- *One Talker / 200 Sentences* and *200 Talkers / 200 Sentences*: not significant ($t(19)=1.87$, $p=0.076$)
- *One Talker / One Sentence* and *200 Talkers / 200 Sentences*: significant ($t(19)=3.75$, $p<.01$)

- **Talker Variability does affect TCE magnitudes, shown by the decreased magnitudes in conditions with increased variability.**




## DISCUSSION

- The results show that TCEs magnitudes are larger following context spoken by a single talker, and smaller following context spoken by different talkers.

- Assgari & Stilp (2015) saw a similar pattern of results in their study regarding Talker Variability and SCEs. But, SCEs their in one-talker conditions were equal and here TCEs in one-talker conditions were marginally but not significantly different. So, the pattern is not perfectly identical.
- It is possible listeners don't just adjust to talker variability, but sentence structure as well.

- Bosker & Ghitza (2018) proposed an idea about what is happening in the brain when listeners are adjusting to speech rate.
  - Neural entrainment – populations of cortical neurons lock onto the speaking rate of speech and fire in response to its syllables.
  - When cortical theta oscillations can optimally track the rate of speech, there are larger context effects.
- Context sentence and speaking rate variability may disrupt this process.

- Acoustic properties of everyday speech are incredibly variable, so it's important to ask if research reflects reality.
- The standard way of measuring contrast effects is by using a single sentence with varying rates of speech. But, in real life, people don't hear a singular sentence back-to-back.
- The current study used repetitive and unique context sentences within and across talkers to better mimic listening situations in the real world.

- A follow-up study with the three original conditions and a new condition of *200 Talkers / One Sentence* would reveal how influential Talker Variability is in speech perception and answer whether sentence structure is an impactful form of context.

## REFERENCES

1. Anwyl-Irvine, A. L., Massonnié, J., Flitton, A., Kirkham, N., & Evershed, J. K. (2020). Gorilla in our midst: An online behavioral experiment builder. *Behavior Research Methods*, *52*(1), 388–407. https://doi.org/10.3758/s13428-019-01237-x
2. Assgari, A. A., & Stilp, C. E. (2015). Talker information influences spectral contrast effects in speech categorization. *The Journal of the Acoustical Society of America*, *138(5)*, 3023-3032.
3. Boersma, P. & Weenink, D. (2020). Pratt: doing phonetics by computer [Computer program]. Version 6.1.24, retrieved 29 September 2020 from http://www.praat.org/
4. Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, *33*(8), 955-967.
5. Ladefoged, P., & Broadbent, D. E. (1957). Information conveyed by vowels. *The Journal of the Acoustical society of America*, *29*(1), 98-104.
6. Stilp, C. E., & Theodore, R. M. (2020). Talker normalization is mediated by structured indexical information. *Attention, Perception, & Psychophysics*, *82*(5), 2237–2243. https://doi.org/10.3758/s13414-020-01971-x
7. Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074-1095.
8. Winn, M. B. (2020). Manipulation of voice onset time in speech stimuli: A tutorial and flexible Praat script. *The Journal of the Acoustical Society of America*, *147*(2), 852–866. https://doi.org/10.1121/10.0000692