How does structured variability help talker adaption?

Claire Suh, Meghan Clayards, Hye-Young Bang
ARIA Research Internship

Introduction

Multiple factors such as speaking style, emotional state, physical condition, and individual talker differences cause speech to vary from moment to moment, and person to person. Despite this, people are able to understand speech with accuracy and speed.

Professor Clayards and Hye-Young Bang have shown that talkers identify consistent correlations across seemingly unrelated sounds. They argue that these correlations are systematic differences between different talkers' speech.

Taking this premise, we are investigating the assertion that people identify these systematic differences based on how hyper-articulated a talker's speech is and that they use these differences to adapt to a novel voice. In other words, we wanted to see if people determine how "careful" or "clear" a novel talker's speech is and if they use that knowledge to learn said talker's voice quickly.

Questions

When listeners hear hyper- or hypo-articulation in a talker's speech in /p, t/-initial words:

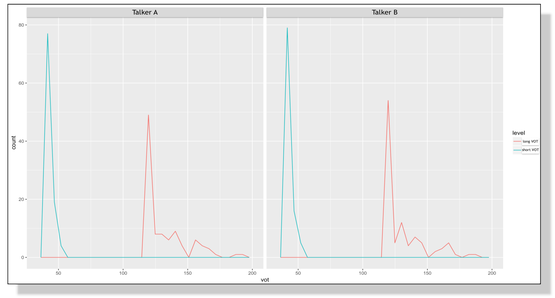
- Can they only generalize within a manner category (within the stop category)?
  - They assume that /k/-initial words follow the type of articulation present in /p, t/-initial words but don't generalize to /s/-duration and /s/-CoG.
  - Participants trained on /p, t/-initial words and tested on /k/-initial prefer the VOT length that matches their training.
- Can they generalize across manner categories only within a temporal domain?
  - They can generalize to /k/-initial words and /s/-duration, but fail to generalize to /s/-CoG.
  - Participants trained on /p, t/-initial words and tested on /k, s/-initial words prefer the VOT length and /s/-duration that matches their training.
- Can they generalize across manner categories and across both temporal and spectral cues?
  - They can generalize to /k/-initial words, /s/-duration, and /s/-CoG.
  - Participants trained on /p, t/-initial words and tested on /k, s/-initial words prefer the VOT length, /s/-duration, and /s/-CoG that matches their training.

Process

- Hyper-/hypo-articulation will be measured by the following factors:
  - VOT (voice onset time) length for stop consonants
  - /s/ duration (length of /s/)
  - /s/ center of gravity (/s/ height in spectrum)
- Stimuli were recorded by two female native Canadian English talkers with a Zoom Handy Recorder H4n in a soundproof room. Each talker recorded:
  - 5 /p/-initial words and 5 /t/-initial words for the familiarization phase
  - 50 /p/-initial words and 50 /t/-initial words for the training phases
  - 20 /k/ words for the /k/ test trials
  - 40 /s/ words for the /s/ test trials
- The words were annotated and run through PRAAT scripts that created the hyper- and hypo-articulated stimuli
  - Hypo- script for /p, t, k/, /s/-dur: Decrease VOT to 40% of original VOT; all VOTs must be at least 40ms
  - Hyper- script for /s/-dur: Increase VOT to 160% of original VOT; all VOTs must be at least 120ms
  - Hyper- script for /p, t, k/: Increase VOT to 200% of original VOT; all VOTs must be at least 120ms
  - Hyper- script for /s/-cog: Filter (one formant): 15,000Hz, 3,000Hz + Filter (pre-emphasis): 50Hz
  - Hypo- script for /s/-cog: Filter (one formant): 6,000Hz, 5,000Hz
- The distributions of the VOTs and CoGs of both talkers were checked by creating histograms in R

Method

- The participant will listen to two talkers, one who has hyper-articulated speech and the other who has hypo-articulated, who are each associated with a picture and a name.
- In the beginning of the experiment, a stimulus is presented with its corresponding picture and name to familiarize the participant to the two talkers.
- After the familiarization phase, there will be ten blocks of the experiment, each consisting of a training phase and a test phase.
- In the training phase, a single training trial consists of a /p, t/ stimulus being played and the participant choosing which talker is speaking the stimulus.
  - The participant receives feedback for talker choice.
  - There are ten training trials per phase.
- In the test phase, the participant will hear a pair of hyper-/hypo-articulated stimuli and must decide which of the pair sounds like one of the talkers. There are three different tests:
  - /k/ (VOT): learn /p, t/ stimuli, test on /k/ stimuli pair
  - /s/-dur: learn /p, t/ stimuli, test on /s/ stimuli pair
  - /s/-cog: learn /p, t/ stimuli, test on /s/ stimuli pair
  - The participant will always be tested on the same talker.
  - Feedback will not be given
  - There are three test trials per phase, with one type of test trial per phase



Histogram comparing training /p, t/ long VOT and short VOT stimuli

Katie

Emma

Pictures and names associated with the two talker voices

References

Theodore, R. M., Myers, E. B., & Lomibao, J. A. (2015). Talker-specific influences on phonetic category structure. *The Journal of the Acoustical Society of America, 138*(2), 1068-1078.

Chodroff, Eleanor, et al. "Structured variability in acoustic realization: A corpus study of voice onset time in American English stops." *The Scottish Consortium for ICPhS* (2015).