Auditory feedback is used for both online and adaptive control of timing in speech

Robin Karlin[1], Chris Naber[1], and Benjamin Parrell[1,2]

[1]Waisman Center, UW-Madison; [2]Department of Communication Sciences and Disorders, UW-Madison

Get preprint

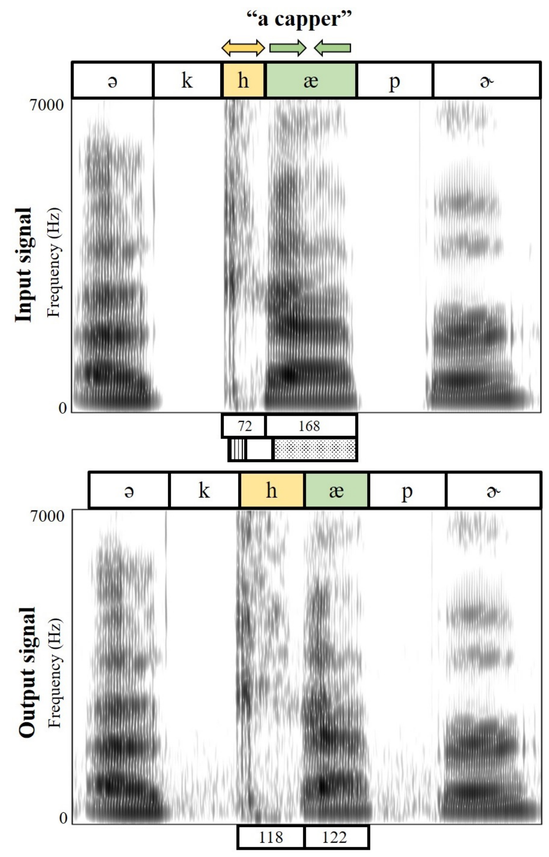
Introduction

The main questions addressed in this study are: Does adaptation differ between relative timing and absolute timing? Is speech timing assessed and controlled via the duration of individual segments, or proportional timing within a larger unit?

- Auditory feedback is used to update forward models of speech motor commands, both in the spectral[3;12;5;2] and the temporal domains
- It is unclear what the limits on temporal adaptation are in speech; some studies have found that speakers adapt the timing of syllable onsets[9], while others have found that only segments in the rime show adaptive behavior[10;4]
- It is also unclear how speech timing is processed and controlled:
  - How does speech timing map to the divide between absolute timing of single events vs. the relative timing of multiple events[14;7;6]?
  - Do speakers attend to the proportional duration of segments in higher prosodic units, or the ms duration of any individual segment[8;11]?

Methodology

- Four target words:
  - Relative timing (VOT): "a **c***a*pper" "a **g***a*pper"*
  - Absolute timing (fricative duration): "a **s***a*pper" "a **z***a*pper"
  - Stressed vowel in each word also absolute timing candidate
- Consonant targets were lengthened and vowel targets were shortened using Audapter[1]
- 20 participants repeated each target word 150 times, in four phases: Baseline (30 trials), Ramp (30), Hold (60), Washout (30)
- Compared[13] production in hold, early washout, and late washout phases to baseline productions



Results: Adaptation in duration of vowel; proportion of consonant target

Consonant target: no adaptation

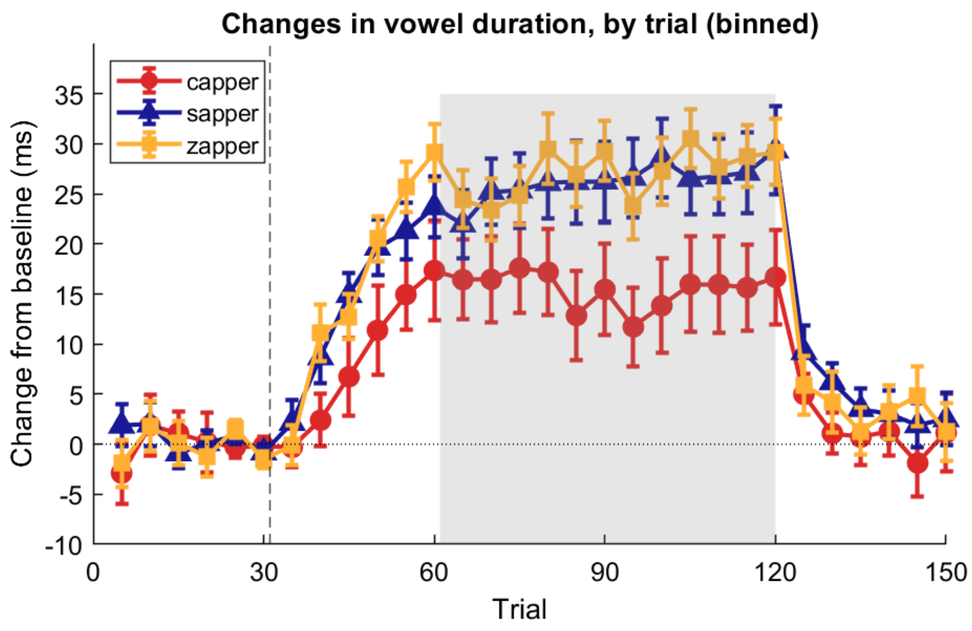
- Participants did not adapt their productions of consonant targets (hold not different from baseline, p = 0.51)
- No difference in adaptation between words ($\chi^2(2) = 2.59$, p = 0.27)

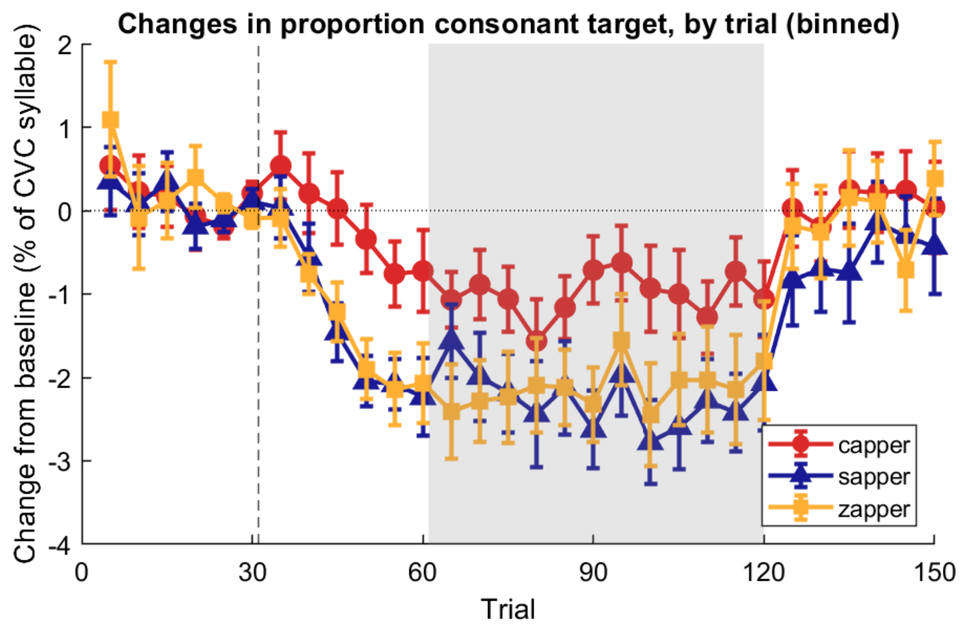
Vowel target: compensation and adaptation

- Participants adapted their productions of vowel targets ($\chi^2(3) = 801.11$, $p < 0.0001$)
- Vowels significantly longer during hold phase by ~80% of perturbation; also remained significantly longer during early washout

Adaptation in consonant target as proportion of CVC syllable

- Consonant target occupies lower proportion of first syllable during hold phase ($\chi^2(3) = 127.24$, $p < 0.0001$), effectively shortening it relative to the rest of the syllable
- Indicates that vowel length difference was not just a speech rate change (initial "a" also unchanged from baseline to hold, $p > 0.17$)
- Suggests that people may attend to proportional duration within larger units rather than absolute duration of particular segments





Discussion

- Speakers adapted the duration of segments only in the rime, aligning with recent research[10]
- Study lends support to the idea that speakers attend to proportional durations, not raw durations[8;11]
  - Difference in response magnitude between hold and early washout phases suggests both compensation and adaptation at work in hold, but how?
  - If using proportions: hear that consonant was longer than expected, compensate by increasing the vowel duration
  - Also hear that vowel shorter than expected and adapt; learning carries over into early washout
- Additional research needed:
  - Why are vowels more amenable to adaptation? Perhaps stiffness, or different mapping of neural timing types[10]
  - How does proportional control interact with absolute vs. relative timing?

References

[1] Shanqing Cai, Satrajit S Ghosh, Frank H Guenther, and Joseph S Perkell. Adaptive auditory feedback control of the production of formant trajectories in the Mandarin triphthong /iau/and its pattern of generalization. *The Journal of the Acoustical Society of America*, 128(4):2033–2048, 2010.

[2] Elizabeth D Casserly. Speaker compensation for local perturbation of fricative acoustic feedback. *The Journal of the Acoustical Society of America*, 129(4):2181–2190, 2011.

[3] Jeffrey L Elman. Effects of frequency-shifted feedback on the pitch of vocal productions. *The Journal of the Acoustical Society of America*, 70(1):45–50, 1981.

[4] Mareike Floegel, Susanne Fuchs, and Christian A Kell. Differential contributions of the two cerebral hemispheres to temporal and spectral speech feedback control. *Nature Communications*, 11(1):1–12, 2020.

[5] John F Houde and Michael I Jordan. Sensorimotor adaptation of speech I: Compensation and adaptation. *Journal of Speech, Language, and Hearing Research*, 45:295–310, 2002.

[6] Richard B Ivry and Steven W Keele. Timing functions of the cerebellum. *Journal of cognitive neuroscience*, 1(2):136–152, 1989.

[7] Richard B Ivry, Rebecca M Spencer, Howard N Zelaznik, and Jörn Diedrichsen. The cerebellum and event timing. *Annals of the new York Academy of Sciences*, 978(1):302–317, 2002.

[8] Rachel H Kessinger and Sheila E Blumstein. Effects of speaking rate on voice-onset time and vowel production: Some implications for perception studies. *Journal of Phonetics*, 26(2):117–128, 1998.

[9] Takashi Mitsuya, Ewen N MacDonald, and Kevin G Munhall. Temporal control and compensation for perturbed voicing feedback. *The Journal of the Acoustical Society of America*, 135(5):2986–2994, 2014.

[10] Miriam Oschkinat and Philip Hoole. Compensation to real-time temporal auditory feedback perturbation depends on syllable position. *The Journal of the Acoustical Society of America*, 148(3):1478–1495, 2020.

[11] Robert F Port and Jonathan Dalby. Consonant/vowel ratio as a cue for voicing in English. *Perception & Psychophysics*, 32(2):141–152, 1982.

[12] David W Purcell and Kevin G Munhall. Adaptive control of vowel formant frequency: Evidence from real-time formant manipulation. *The Journal of the Acoustical Society of America*, 120(2):966–977, 2006.

[13] R Core Team. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2019.

[14] Sundeep Teki, Manon Grube, Sukhbinder Kumar, and Timothy D Griffiths. Distinct neural substrates of duration-based and beat-based auditory timing. *Journal of Neuroscience*, 31(10):3805–3812, 2011.

Contact

rkarlin@wisc.edu smac.waisman.wisc.edu @MotorSpeech