Vowel Duration Is Encoded Relative to the Contextual Speech Rate Throughout Cortical Auditory Processing

Mara Wolter (m.wolter@uni-tuebingen.de)
Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25,
Graduate Training Centre of Neuroscience, University of Tübingen, Otfried-Müller-Str. 27,
72076 Tübingen, Germany

Andrey Zyryanov (andrey.zyryanov@uni-tuebingen.de)
Centre for Integrative Neuroscience, University of Tübingen, Otfried-Müller-Str. 25,
Graduate Training Centre of Neuroscience, University of Tübingen, Otfried-Müller-Str. 27,
72076 Tübingen, Germany

Yulia Oganian (yulia.oganian@uni-tuebingen.de)
Centre for Integrative Neuroscience, Otfried-Müller-Str. 25, 72076 Tübingen, Germany

Abstract

Across languages, listeners perceive durational features in speech relative to the syllable rate of the surrounding context. Despite the importance of this mechanism, it remains unclear how the cortex normalizes sound durations during speech perception. Here we used MEG to investigate rate normalization for vowel duration in German, where the contextual speech rate can decisively shift listeners' word perception (e.g., *satt* – ˈzat – "full" vs. *Saat* – ˈzaːt – "seed"). We found that rate influences evoked activity before, during and after the target vowel, across auditory, motor and inferior frontal regions. Vowel duration effects from ~130 ms post vowel offset always co-occurred with rate effects. This spatiotemporal overlap suggests that vowel duration is encoded relative to contextual speech rate throughout cortical processing, and that contextual rate normalization is an integral part of duration encoding.

Keywords: Rate normalization, auditory cortex, speech processing, magnetoencephalography

Introduction

Durational cues in speech are perceived relative to the contextual speech rate: a fixed duration is perceived as longer when heard in a fast context (Maslowski, Meyer and Bosker, 2019; Miller, 1981; Summerfield, 1981). This robust behavioral effect offers a window into the neural mechanisms behind contextual normalization of durational cues.

In the auditory speech cortex, located on the transverse temporal plane (TTP) in the Sylvian fissure and on the superior temporal gyrus (STG), evoked response latencies increase and magnitudes decrease for longer sounds (Inouchi et al., 2008; Toscano et al., 2021). Notably, amplitude differences around 100 ms post stimulus onset have been linked to perceptual categorization, with identical durations eliciting a lower response magnitude when they were perceived as long as compared to short (Ou & Yu, 2022).

At the same time, the auditory speech cortex tracks speech rate by phase-locking to the rising acoustic edges in the speech envelope (Oganian & Chang, 2019), and auditory–motor phase-locking at the syllable rate increases during listening within natural rate ranges (Assaneo & Poeppel, 2019). Moreover, Kösem et al. (2018) found that phase-locking at the speech rate frequency in right temporal and inferior frontal regions predicted individual perceptual biases in rate normalization for vowels.

Yet, how contextual speech rate and durational cues interact at the neural level remains unknown. We hypothesize that both factors shape auditory evoked responses following vowel offset. Specifically, in line with behavioral findings, we predict that fast contextual rate will modulate neural responses in the same direction as a longer vowel duration, suggesting that vowel duration is encoded relative to contextual rate where effects overlap in space and time. We expect that vowel duration is initially processed in absolute terms, and normalized at later stages of processing.

Methods

Thirty participants (14m/15f/1d) completed the MEG experiment and structural MRI scan. We presented three target word pairs (*satt*–*Saat*, *Bann*–*Bahn*, *Rum*–*Ruhm*) along 7-step vowel duration continua. The continua were individually adjusted for each participant, based on a preceding staircase procedure. Targets were preceded by a 5-syllable fast or slow precursor sentence (7/2.5 syl/s; Fig. 1A). After each trial, participants indicated which word they perceived. In total, 840 trials were recorded (20 per word pair x vowel duration x rate combination).

Source reconstruction on band-pass filtered (0.5 - 40 Hz) and cleaned data was performed using the ico-4 surface with fixed source locations. We modeled source-level responses time-locked to vowel offset with a linear regression with predictors for precursor rate, target word, vowel duration and all interactions. Analyses were performed both on whole-brain data and in a set of pre-defined anatomical regions of interest (e.g., TTP, Destrieux et al. 2010).

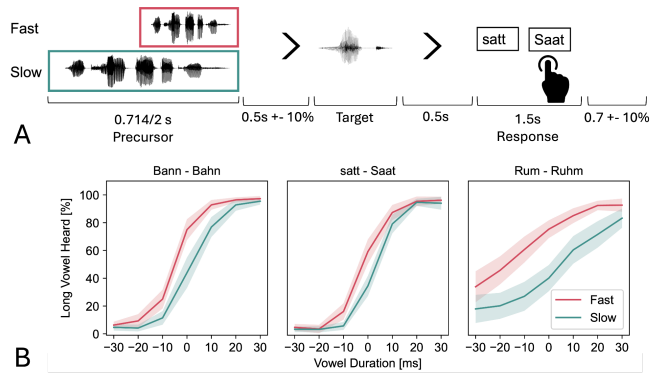

Figure 1: A: Paradigm. Each trial consisted of a precursor sentence, followed by the target word and a response screen. B: Behavioral responses (95% confidence interval) as a function of vowel duration, relative to the midpoint of the duration continuum.

Results

As expected, we observed a robust behavioral effect of the contextual rate: The fast-rate context led to an average 13.6% ($p<0.001$) increase in long vowel responses compared to the slow-rate context (Figure 1B). Neurally, we observed a significant bilateral effect of speech rate before and during presentation of the target words, around auditory and speech motor (ventral central sulcus) areas (Dichter et al., 2018). We observed a brief right-lateralized effect of vowel duration around 70 ms, likely due to latency differences (Figure 2A). The effect of vowel duration re-emerged between 180–390 ms in the right TTP, posterior STG, and mid superior temporal sulcus, with significantly lower response magnitudes for long vowel durations. Additionally, longer vowel durations evoked a significantly lower response magnitude around bilateral inferior frontal gyri, ~130-240 ms and ~190-270 ms post vowel offset, respectively. (Figure 2B). Notably, duration effects post-vowel offset always co-occurred with rate effects, resulting in a shift in neural responses with faster rates in line with the behavioral normalization effect.

Summary and Conclusion

Our findings show that vowel duration is encoded relative to contextual speech rate throughout cortical processing. This suggests that speech rate normalization is not a separate, downstream adjustment but an integral part of perceptual encoding, including early auditory responses.

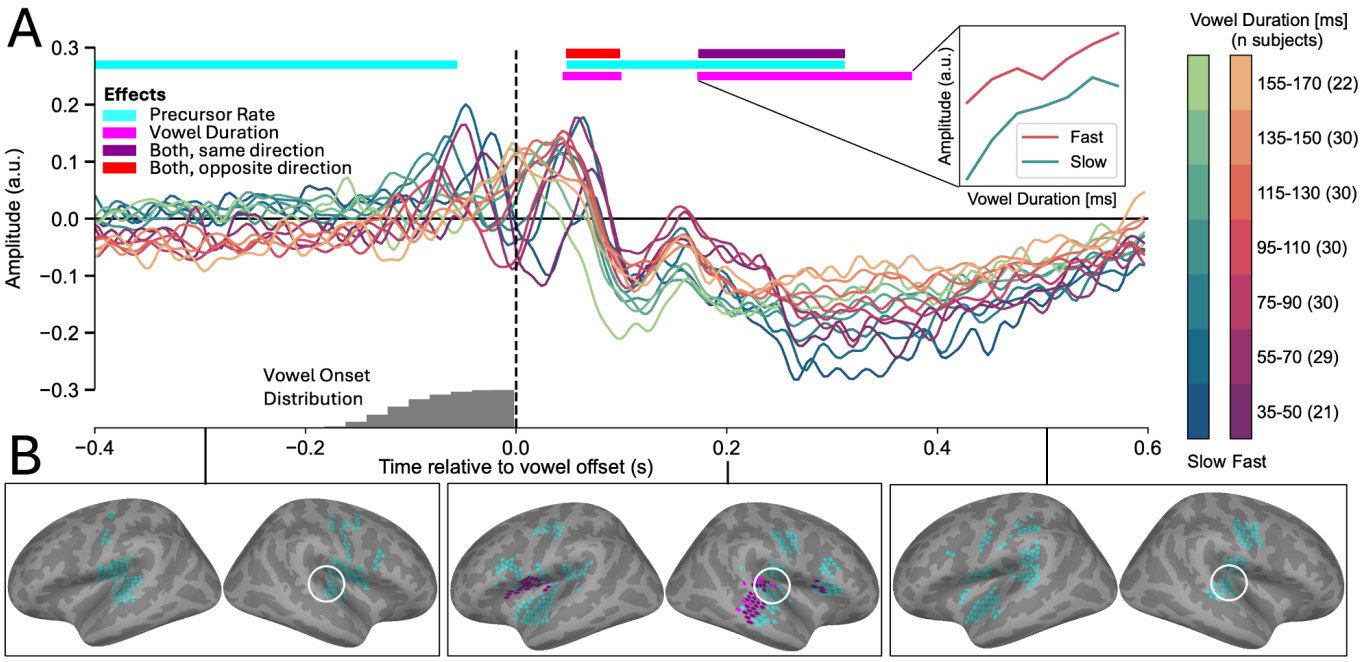

Figure 2: A: Evoked responses in the right transverse temporal plane (TTP, white circles in panel B). Bars mark significant clusters (individual $p<0.1$, cluster $p<0.05$). Trials with slow-rate precursors in blue-green; fast-rate in purple-orange. B: Spatial distribution of rate and duration effects at three selected timepoints.

Acknowledgements

The authors thank the International Max Planck Research School for the Mechanisms of Mental Function and Dysfunction (IMPRS-MMFD) for supporting Mara Wolter and Andrey Zyryanov.

References

Assaneo, M. F., & Poeppel, D. (2018). The coupling between auditory and motor cortices is rate-restricted: Evidence for an intrinsic speech-motor rhythm. *Science advances*, *4*(2), eaao3842.

Destrieux, C., Fischl, B., Dale, A., & Halgren, E. (2010). Automatic parcellation of human cortical gyri and sulci using standard anatomical nomenclature. *Neuroimage*, *53*(1), 1-15.

Dichter, B. K., Breshears, J. D., Leonard, M. K., & Chang, E. F. (2018). The control of vocal pitch in human laryngeal motor cortex. *Cell*, *174*(1), 21-31.

Inouchi, M., Kubota, M., Ohta, K., Matsushima, E., Ferrari, P., & Scovel, T. (2008). Neuromagnetic mismatch field (MMF) dependence on the auditory temporal integration window and the existence of categorical boundaries: Comparisons between dissyllabic words and their equivalent tones. *Brain research*, *1232*, 155-162.

Kösem, A., Bosker, H. R., Takashima, A., Meyer, A., Jensen, O., & Hagoort, P. (2018). Neural entrainment determines the words we hear. *Current Biology*, *28*(18), 2867-2875.

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 128.

Miller, J. L. (1981). Some effects of speaking rate on phonetic perception. *Phonetica*, *38*(1-3), 159-180.

Oganian, Y., & Chang, E. F. (2019). A speech envelope landmark for syllable encoding in human superior temporal gyrus. *Science advances*, *5*(11), eaay6279.

Ou, J., & Yu, A. C. (2022). Neural correlates of individual differences in speech categorisation: Evidence from subcortical, cortical, and behavioural measures. *Language, Cognition and Neuroscience*, *37*(3), 269-284.

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074.

Toscano, J. C., Anderson, N. D., Fabiani, M., Gratton, G., & Garnsey, S. M. (2018). The time-course of cortical responses to speech revealed by fast optical imaging. *Brain and Language*, *184*, 32-42.