Listening under cognitive load makes speech sound fast

Hans Rutger Bosker [1]*, Eva Reinisch* [2]*, and Matthias Sjerps* [3]

[1] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands
[2] Institute of Phonetics and Speech Processing, Ludwig Maximilian University Munich, Germany
[3] Department of Linguistics, University of California, Berkeley, CA, USA

HansRutger.Bosker@mpi.nl; evarei@phonetik.uni-muenchen.de; m.j.sjerps@gmail.com

Index Terms: cognitive load, rate normalization, speech rate, speech perception

1. Introduction

Listeners interpret local temporal cues (e.g., vowel durations) relative to the surrounding speech rate. For instance, an ambiguous Dutch vowel midway between short /ɑ/ and long /a:/ may be perceived as long /a:/ when presented in a fast context, but as short /ɑ/ in a slow context [1]. It is widely assumed that this process known as rate normalization is an early general auditory process [1, 2], and as such would operate independent from other higher level influences such as attention. However, when the perceptual system is taxed by the concurrent execution of another task, the encoding of the incoming speech signal is known to be negatively affected [3]. Therefore, listening to, for example, fast speech under cognitive load may result in impoverished encoding of the fast speech rate, *reducing* the effect that a fast context may have on the perception of subsequent speech (i.e., a reduction of the rate effect; cf. [4]). Alternatively, an increase in cognitive load has been shown to speed up time perception (the "shrinking of time", [5]), potentially *increasing* the perceived rate of concurrent speech. This argument has, for instance, been used to explain why foreign-accented speech sounds faster than native speech [6].

Here we attempt to distinguish between these alternatives by testing Dutch /ɑ/-/a:/ categorization as a function of (1) the rate of the preceding carrier sentence and (2) the difficulty of a dual task (visual search) performed during carrier presentation.

2. Method

Data from 29 participants with normal hearing and vision were collected. Dutch speech materials were adopted from [1], including a semantically neutral carrier sentence in a fast (793 ms) and a slow version (1648 ms), that ended in a minimal word pair. The vowel in the sentence-final target word was spectrally ambiguous between /ɑ/ and /a:/ and was presented at three different durations in the ambiguous range (V1 = relatively long, V3 = relatively short; see [1]).

The dual task (visual search) was adopted from [3]. Visual displays consisted of an equal number of black squares, black triangles, red triangles, red diamonds, and on half of the trials one oddball: a black diamond. Low and high cognitive load conditions (4x4 vs. 13x13 grids) were blocked (order counter-balanced across participants; total 432 trials).

A trial started with the onset of the visual grid. After 250 ms of silence, the carrier sentence was played. At carrier offset, the visual grid was replaced by a response screen and
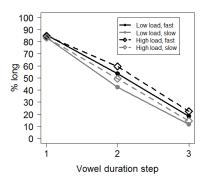

Figure 1: *Average percentage long vowel responses split by load condition and carrier rate.*

the target word was presented. Note that the visual search task was only performed during carrier presentation, not during the target word. Participants first indicated by button press which word they had heard sentence-finally, and then indicated whether or not they had seen the oddball in the visual display.

3. Results

Visual search accuracy differed across cognitive load conditions (low: 96%; high: 68%) suggesting that the load manipulation was effective. Speech target categorization functions are displayed in Figure 1. Statistical analyses using GLM models [7] show main effects for the two Vowel Step comparisons (V2 vs. V1, β = 2.262, p < 0.001; V2 vs. V3, β = 2.344, p < 0.001), Carrier Rate (β = 0.538, p < 0.001), and Load (β = 0.411, p = 0.004). No interaction between Load and Rate was found, but an interaction between Rate and Vowel Step indicated a larger effect of Rate for Vowel Step 3 (β = 0.451, p = 0.001).

4. Discussion

We tested whether the presumed early auditory process of rate normalization may be affected by cognitive load. Results indicated that cognitive load did *not* modulate rate normalization. Rather, a main effect of load indicated that the execution of a concurrent secondary task led to a higher proportion of long vowel responses. This suggests that under cognitive load, the carrier sentences were perceived as relatively fast. This main effect could be explained by a model for contextual influences on duration judgments that involves 1) an automatic mechanism [1, 2] unaffected by attention; and 2) an attention-dependent mechanism, accounting for higher level influences (cf. [6]) such as cognitive load.

5. References

[1] E. Reinisch and M. J. Sjerps, "The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context," *Journal of Phonetics,* vol. 41, pp. 101-116, 2013.

[2] M. J. Sjerps and E. Reinisch, "Divide and conquer: How perceptual contrast sensitivity and perceptual learning cooperate in reducing input variation in speech perception," *Journal of Experimental Psychology: Human Perception and Performance,* vol. 41, pp. 710-722, 2015.

[3] S. L. Mattys and L. Wiget, "Effects of cognitive load on speech recognition," *Journal of Memory and Language,* vol. 65, pp. 145-160, 2011.

[4] M. A. Pitt, C. Szostak, and L. Dilley, "Rate dependent speech processing can be speech-specific: Evidence from the perceptual disappearance of words under changes in context speech rate," *Attention, Perception, & Psychophysics,* pp. 1-12, 2015.

[5] R. A. Block, P. A. Hancock, and D. Zakay, "How cognitive load affects duration judgments: A meta-analytic review," *Acta Psychologica,* vol. 134, pp. 330-343, 2010.

[6] H. R. Bosker and E. Reinisch, "Normalization for speechrate in native and nonnative speech," in *Proceedings of the 18th International Congress of Phonetic Sciences 2015 [ICPhS XVIII], Glasgow*, 2015.

[7] H. Quené and H. Van den Bergh, "Examples of mixed-effects modeling with crossed random effects and with binomial data," *Journal of Memory and Language,* vol. 59, pp. 413-425, 2008.