Selective attention to a specific talker does not change the effect of surrounding acoustic context

Hans Rutger Bosker[a,b], Eva Reinisch[c], Matthias Sjerps[a,b]

[a]*Max Planck Institute for Psycholinguistics, The Netherlands*

[b]*Donders Institute for Brain, Cognition and Behaviour, Radboud University, Netherlands*

[c]*Institute of Phonetics and Speech Processing, LMU Munich, Germany*

Spoken sentences contain considerable prosodic variation, for instance in their speech rate [1]. One mechanism by which the listener can overcome such variation is by interpreting the durations of speech sounds relative to the surrounding speech rate. Indeed, in a fast context, a durationally ambiguous sound is perceived as longer than in a slow context [2].

In abstractionist models of spoken word comprehension, this process – known as rate normalization – affects pre-lexical representations before abstract phonological representations are accessed [3]. A recent study [4] provided support for such an early perceptual locus of rate normalization. In that study, participants performed a visual search task that induced high (large grid) vs. low (small grid) cognitive load, while listening to fast and slow context sentences. Context sentences were followed by durationally ambiguous targets. Fast sentences were shown to bias target perception towards more ‘long’ target segments than slow contexts. Critically, changes in cognitive load did *not* modulate this rate effect. These findings support a model in which normalization processes arise early during perceptual processing; too early to be affected by attentional modulation.

The present study further evaluated the cognitive locus of normalization processes by testing the influence of another form of attention: auditory stream segregation. Specifically, if listeners are presented with a fast and a slow talker *at the same time but in different ears*, does explicitly attending to one or the other stream influence target perception? The aforementioned model [4] predicts that selective attention should not influence target perception, since normalization processes should be robust against changes in attention allocation. Alternatively, if attention does modulate normalization processes, two participants, one attending to fast, the other to slow speech, should show different perception.

Dutch participants (Expt 1: N=32; Expt 2: N=16; Expt 3: N=16) were presented with 200 fast and slow context sentences of various lengths, followed by a target duration continuum ambiguous between, e.g., short target “geven” /ˈxevə/ *give* vs. long target “gegeven” /**xə**ˈxevə/ *given* (i.e., 20 target pairs differing presence/absence of unstressed syllable /**xə**-/).

Critically, in Experiment 1, participants heard two talkers simultaneously (talker and location counter-balanced across participants), one (relatively long) sentence at a fast rate, and one (half as long) sentence at a slow rate (rate varied within participants). Context sentences were followed by ambiguous targets from yet another talker (Fig. 1). Half of the participants was instructed to attend to talker A, while the other half attended to talker B. Thus, participants heard identical auditory stimuli, but varied in which talker they attended to.

Debriefing questionnaires and transcriptions of attended talkers in filler trials confirmed that participants successfully attended to one talker, and ignored the other. Nevertheless, no effect of attended rate was found (Fig. 2; p>.9), indicating that modulation of attention did not influence participants’ rate normalization. Control experiments showed that it was possible to obtain rate effects with single talker contexts that were either talker-incongruent (Expt 2) or talker-congruent (Expt 3) with the following target (Fig. 1). In both of these experiments, there was a higher proportion of long target responses following a fast context (Fig. 2). This shows that contextual rate affected the perception of syllabic duration and that talker-congruency with the target did not change the effect. Therefore, in line with [4], the current experiments suggest that normalization processes arise early in perception, and are robust against changes in attention.

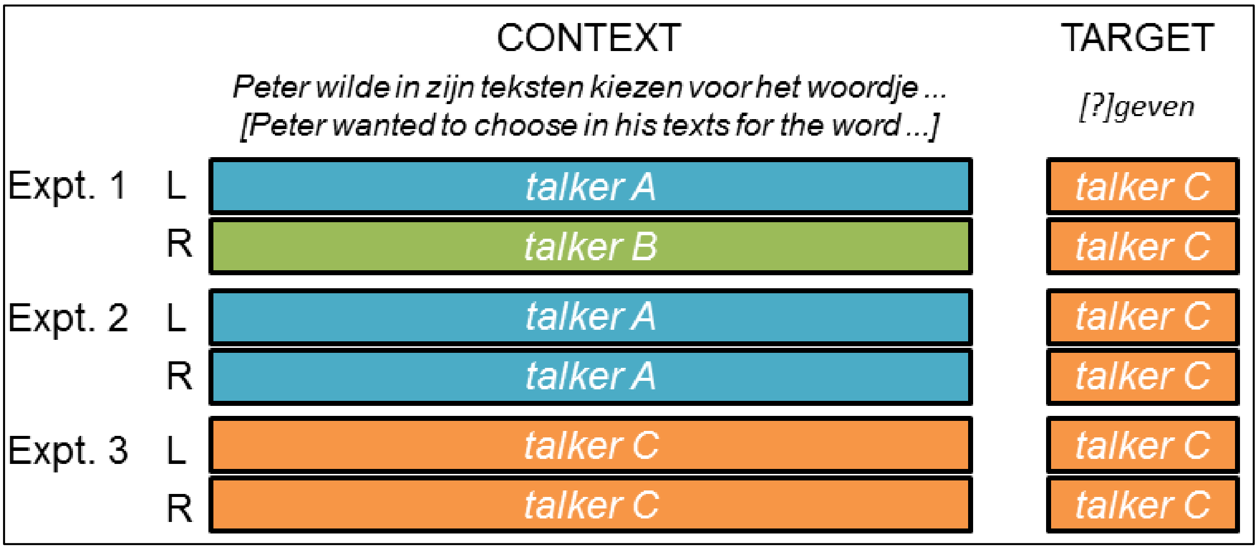

Figure 1. *Design of the three experiments.*

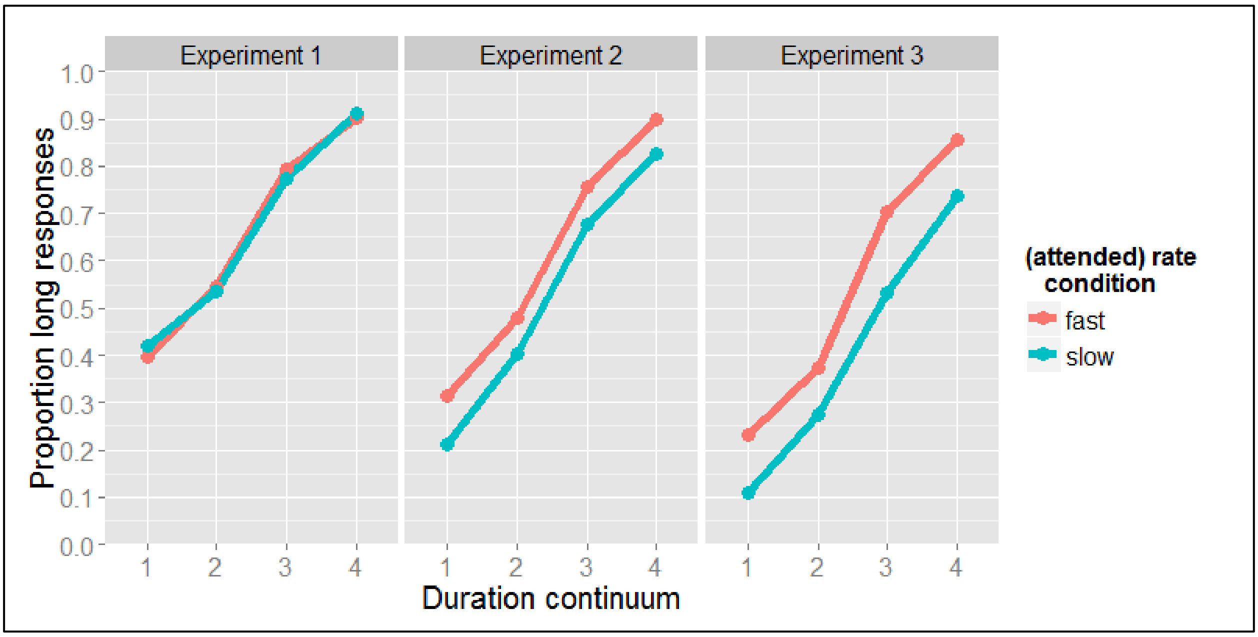

Figure 2. *Proportion of long target responses, split by rate, separately for each experiment.*

REFERENCES

[1] J. L. Miller, F. Grosjean, and C. Lomanto, "Articulation rate and its variability in spontaneous speech: A reanalysis and some implications," *Phonetica*, vol. 41, pp. 215-225, 1984.

[2] 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.

[3] J. M. McQueen and A. Cutler, "Spoken word access processes: An introduction," *Language and Cognitive Processes*, vol. 16, pp. 469-490, 2001.

[4] H. R. Bosker, E. Reinisch, and M. J. Sjerps, "Cognitive load makes speech sound fast but does not modulate acoustic context effects," *Journal of Memory and Language*, vol. 94, pp. 166-176, 2017.