When speech cues are not integrated immediately: evidence from the global speech rate effect

Merel Maslowski (*Royal Dutch Kentalis*) and Antje Meyer *(Max Planck Institute for Psycholinguistics)* and Hans Rutger Bosker *(Max Planck Institute for Psycholinguistics).*
HansRutger.Bosker@mpi.nl

When having a conversation, small differences in the rates at which talkers speak can influence how we understand individual speech sounds. Therefore, in order to successfully comprehend speech, listeners tune in to speech rate information in the surrounding spoken context. For instance, an ambiguous vowel midway between Dutch short /ɑ/ and long /a:/ is more likely to be perceived as long /a:/ when it is preceded by a fast sentence. Conversely, the same vowel is more likely to be perceived as short /ɑ/ when preceded by a slow sentence. That is, the speech rate in the *distal context* contrastively influences the perception of durational cues of a following target word [1].

But listeners have also been shown to take the larger *global context* into account when interpreting vowel duration cues. For instance, Maslowski et al. [2] presented two groups of listeners with vowel continua, manipulating vowel duration in five steps from /ɑ/ to /a:/, embedded in sentences. Talker A's vowel continua were always embedded in sentences, produced by Talker A, with a 'neutral' (=original) speech rate. However, in the High-Rate Group, Talker B's vowel continua were embedded in linearly compressed (multiplied by 1/1.6; fast) sentences from Talker B, while in the Low-Rate Group, Talker B's vowel continua were embedded in expanded sentences (multiplied by 1.6; slow) from Talker B. Critically, the perception of the ambiguous vowels of Talker A depended on the global context: participants in the High-Rate Group perceived Talker A as relatively slow compared to fast Talker B, leading to a lower proportion of long /a:/ reports, compared with the Low-Rate Group.

Recently, empirical findings have indicated different constraints for distal vs. global speech rate effects. For instance, distal rate effects seem to be insensitive to talker identity and can even be induced by one's own voice [3]. In contrast, global rate effects are modulated by talker identity [2] and are not observed when Talker B in the example above is replaced by one's own voice [4]. This has led Bosker et al. [1] to propose a two-stage model of acoustic context effects, suggesting that distal rate effects arise early, while global rate effects occur at a later stage.

The present study involved an eye-tracking experiment with written words on screen to compare the time courses of distal and global rate effects. We adopted the same two-groups paradigm of Maslowski et al. [2] (Fig. 1; each group had n=16; mean age=23, range=20-28), while recording eye fixations to two response options on screen (e.g., *tak* 'branch' vs. *taak* 'task') as participants categorized the ambiguous /ɑ-a:/ target words for a total of 400 trials. A linear mixed model on the categorization data indicated a significant within-groups distal rate effect (more long /a:/ responses in faster speech; Fig. 2A) as well as a significant between-groups global rate effect (fewer long /a:/ responses in Talker A's speech for the High-Rate Group vs. Low-Rate Group; solid lines in Fig. 2A), in line with [2]. Eye fixations to the long /a:/ target word were analyzed by means of a 'Bootstrapped Differences of Timeseries' analysis (BDOTS) and a jackknife procedure. Results indicated that the distal rate effect (more looks to long /a:/ target for the faster rate within groups; Fig. 2B) emerged very early after vowel offset (+308 ms). The global rate effect (fewer looks to long /a:/ target in neutral rate for High-Rate Group; cf. middle two lines in Fig. 2B), however, arose more than 200 ms later than the distal rate effect (+532 ms), suggesting that distal and global context effects involve distinct processing mechanisms.

These findings demonstrate that the context in which words are uttered can systematically change the way speech is perceived, highlighting the complexity of speech comprehension in everyday conversation. Moreover, the results are in line with the two-stage model of acoustic context effects: distal context effects involve early perceptual processes, while global context effects arise at a later stage, involving cognitive adjustments conditioned by higher-level information. The present outcomes are also one of the few findings (cf. [5]) of listeners integrating speech cues relatively late in time, calling for future research disentangling how and why some speech cues are processed earlier than others.

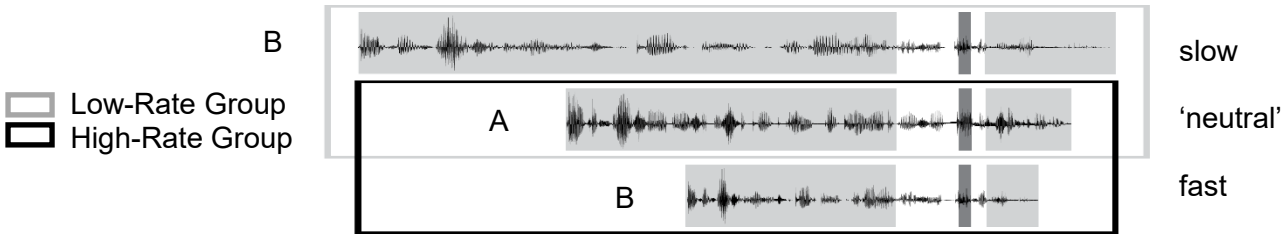

Fig. 1. Experimental design. Targets involved five-step /ɑ-a:/ continua (darkgray) from Talker A and B, embedded in sentences. The High-Rate Group heard Talker A talk at a 'neutral' rate, while Talker B spoke fast. The Low-Rate Group also heard 'neutral' Talker A, but heard Talker B talking slowly.

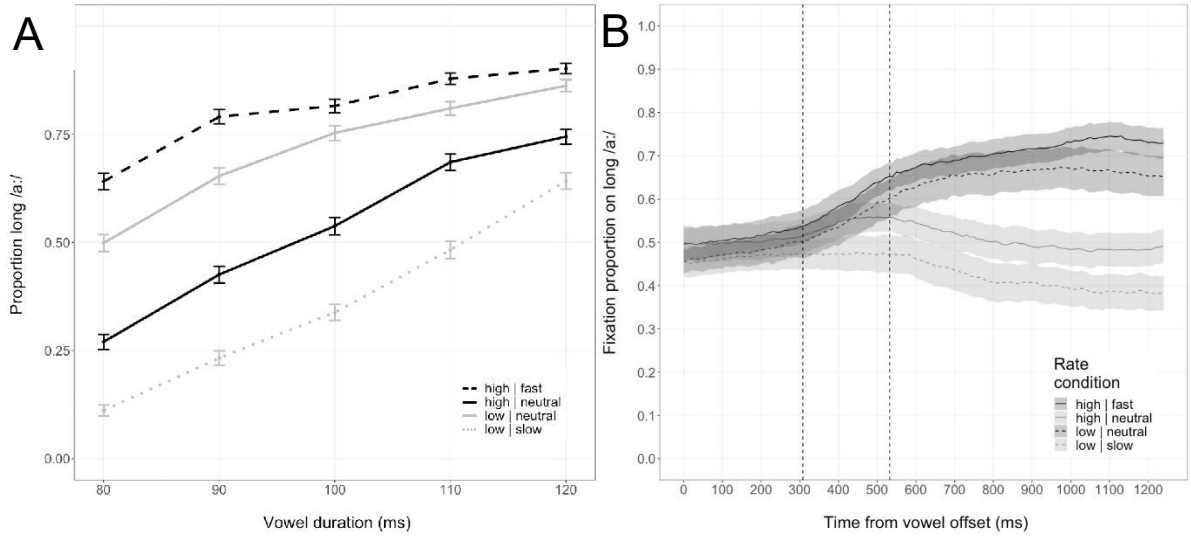

Fig. 2. Experimental results. A. Average categorization data in proportion of long /a:/ responses, with vowel duration continua steps on the x-axis. Color indicates Group (High-Rate Group in black; Low-Rate Group in gray) and line type indicates Rate Condition (dashed for fast, solid for neutral, dotted for slow). The critical comparison for the global rate effect is between the two solid lines, reflecting perception of the neutral speech rate condition in the two groups. Error bars represent the standard error of the mean. **B.** Average fixation proportions to the long /a:/ target word as a function of contextual speech rate, collapsed across vowel durations. Time point 0 is the offset of the target vowel. Line type indicates group (solid for High-Rate Group; dashed for Low-Rate Group). Black lines represent relatively fast speech rates within groups (fast in High-Rate Group; neutral in Low-Rate Group). Gray lines represent relatively slow rates (neutral in High-Rate Group; slow in Low-Rate Group). The two vertical dashed lines indicate the earliest time points at which the distal rate effect (308 ms) and the global rate effect (532 ms) were reliably detectable by the BDOTS analysis.

REFERENCES

[1] 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, doi: 10.1016/j.jml.2016.12.002.

[2] M. Maslowski, A. S. Meyer, and H. R. Bosker, "How the tracking of habitual rate influences speech perception," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 45, no. 1, pp. 128–138, 2019, doi: 10.1037/xlm0000579.

[3] H. R. Bosker, "How our own speech rate influences our perception of others," *Journal of Experimental Psychology: Learning, Memory, and Cognition*, vol. 43, pp. 1225–1238, 2017, doi: 10.1037/xlm0000381.

[4] M. Maslowski, A. S. Meyer, and H. R. Bosker, "Listening to yourself is special: Evidence from global speech rate tracking," *PLOS ONE*, vol. 13, no. 9, p. e0203571, Sep. 2018, doi: 10.1371/journal.pone.0203571.

[5] H. Mitterer, T. Cho, and S. Kim, "How does prosody influence speech categorization?," *Journal of Phonetics*, vol. 54, pp. 68–79, Jan. 2016, doi: 10.1016/j.wocn.2015.09.002.