Eye-tracking the time course of distal and global speech rate effects

Merel Maslowski[a,b*], **Antje S. Meyer**[b,c], **Hans Rutger Bosker**[b,c*]

[a] *Royal Dutch Kentalis, Sint-Michielsgestel, The Netherlands*

[b] *Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands*

[c] *Donders Institute for Brain, Cognition and Behaviour, Radboud University, Nijmegen, The Netherlands*

**Corresponding authors. E-mail addresses: M.Maslowski@kentalis.nl; HansRutger.Bosker@mpi.nl*

Abstract

To comprehend speech sounds, listeners tune in to speech rate information in the proximal (immediately adjacent), distal (non-adjacent), and global context (further removed preceding and following sentences). Effects of global contextual speech rate cues on speech perception have been shown to follow constraints not found for proximal and distal speech rate. Therefore, listeners may process such global cues at distinct time points during word recognition. We conducted a printed-word eye-tracking experiment to compare the time courses of distal and global rate effects. Results indicated that the distal rate effect emerged immediately after target sound presentation, in line with a general-auditory account. The global rate effect, however, arose more than 200 ms later than the distal rate effect, indicating that distal and global context effects involve distinct processing mechanisms. Results are interpreted in a two-stage model of acoustic context effects. This model posits that distal context effects involve very early perceptual processes, while global context effects arise at a later stage, involving cognitive adjustments conditioned by higher-level information.

Keywords: speech rate, rate normalization, distal context, global context, two-stage model, eye-tracking

Public significance statement

When having a conversation, small differences in the rates at which talkers speak can have strong effects on how we understand individual speech sounds. This study shows that such speech rate effects apply to different types of speech rate contexts, defined by their proximity to a temporally ambiguous word. This means that listeners use both the within-sentence speech rate and the speech rate beyond the sentence, from a different talker, to interpret the upcoming speech signal. This study reveals that not only do listeners use both of these different types of speech rate cues, the way in which these cues are processed also differs: Within-sentence rate cues are used in an earlier time window than cues at larger distances. 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.

Introduction

When humans listen to speech, they pick up on many different acoustic cues that contribute to the comprehension of a given word. Specifically, listeners do not only consider the segmental cues of a given word in isolation; instead, they take into account many other acoustic properties of the word itself, its surrounding acoustic context, who produced the word, and even such situational factors as environmental noise and room acoustics. All these factors contribute to achieving the goal of language comprehension, but they may do so at different time points during perceptual processing. In this study, we tested when listeners use different contextual speech rate cues in word recognition.

Listeners track and adapt to temporal information in speech. That is, they use the speech rate context to tune their perception of temporally ambiguous stretches of speech, such as short and long vowels (Bosker, 2017a), consonants (Miller & Baer, 1983), and words (Baese-Berk, Dilley, Henry, Vinke, & Banzina, 2019; Dilley & Pitt, 2010). As such, listeners take into account the surrounding speech rate. Interestingly, listeners have been shown to be sensitive to speech rate in at least three types of context: proximal, distal, and global contexts. The proximal context is defined as the context directly preceding and following an ambiguous stretch of speech to a distance of approximately 250 to 300 ms (Newman & Sawusch, 1996; Reinisch, Jesse, & McQueen, 2011; Sawusch & Newman, 2000; Summerfield, 1981). For instance, Diehl and Walsh (1989) showed that the phonetic category boundary between /b/ (short VOT) and /p/ (long VOT) can be shifted from one phoneme to another by altering the duration of the following vowel; reduction of the duration of the vowel in /ba/ led to a bias towards hearing /pa/.

The distal context is the sentence context beyond the proximal context in both directions, typically the surrounding sentence context (cf. Reinisch et al., 2011). That is, while proximal context is controlled, listeners are more likely to hear an ambiguous Dutch /ɑ–a:/ vowel as short /ɑ/ when the distal context is slow compared to fast (Reinisch & Sjerps, 2013). Conversely, listeners tend to perceive the same ambiguous vowel as long /a:/ when the sentence is fast.

Global speech rate information comes from longer contexts (up to an hour of speech; Baese-Berk et al.,

2014) and multiple talkers, where one talker affects perception of another talker (Maslowski, Meyer, & Bosker, 2018, 2019a). Maslowski et al. (2019a) investigated inter-talker effects of global speech rate on perception of the Dutch vowel contrast between /ɑ/ and /a:/. Two participant groups listened to sentences spoken by two different talkers. In the high-rate group, Talker A always spoke at a neutral speech rate, whereas Talker B had a fast speech rate. In the low-rate group, Talker A's speech rate was again neutral, but Talker B spoke at a slow speech rate. Maslowski et al. found a contrastive effect of global speech rate; if Talker B was fast, neutral Talker A sounded slow, but if Talker B was slow, neutral Talker A sounded fast. This was evidenced by more long /a:/ responses for neutral Talker A in the low-rate group than in the high-rate group. The experimental results were closely replicated in Maslowski et al. (2018) with naturally produced fast and slow speech (cf. their Experiment 3).

In addition to global speech rate affecting the perception of phonemes, there is also evidence that the global speech rate context influences the perception of entire words (Baese-Berk et al., 2014). This concerns the perception of the presence or absence of heavily reduced function words like *our* in phrases such as *Susan said those are (our) black socks*. Coarticulated function words are more often detected when the surrounding speech is perceived as fast compared to slow (known as the 'lexical rate effect'; Dilley & Pitt, 2010; Brown, Dilley, & Tanenhaus, 2012). For instance, Baese-Berk et al. compared three listeners groups who differed only in the average speech rate they heard across different context sentences. They found differences between the groups in their perception of ambiguous targets depending on the average speech rate each group had heard.

In the current study, we focused on the time course of distal and global context effects of surrounding speech rate on vowel categorization (i.e., phonetic boundary shifts). Speech rate cues in the distal context have been suggested to affect the perception of target speech sounds immediately. For instance, Reinisch and Sjerps (2013) investigated the timing of the integration of contextual temporal and spectral cues in a printed-word visual world paradigm. Participants listened to context sentences that were manipulated either temporally (fast vs. slow) or spectrally (high F2 vs. low F2). In these sentences, the proximal context was controlled; that is, each sentence included a fixed 300 ms silent interval preceding and following the target. The authors measured participants' fixations to written words on a screen, to test how the context sentences would affect perception of a following target word with the Dutch /ɑ, a:/ vowel contrast. They found that both spectral and durational cues immediately influenced perception of the target vowel. The effect of fast versus slow distal speech contexts on participants' eye fixations could already be picked up around 300 to 400 ms after vowel onset. Reinisch and Sjerps argued that distal contextual influences happened at a very early stage of processing.

Toscano and McMurray (2015) also investigated how listeners coped with variability in speech rate. They tested the influence of the sentence rate context on VOT, using a visual world paradigm with visual stimuli representing minimal word pairs such as *beach/peach*. They found that speech rate cues immediately modulated the uptake of VOT, as soon as the information was available. Toscano and McMurray's speech rate

effect arose approximately between 300 and 400 ms after target word onset, in corroboration with the distal speech rate effect in Reinisch and Sjerps (2013). For direct comparison with Reinisch and Sjerps, note that Toscano and McMurray measured their effects from target word onset rather than vowel onset. Furthermore, Toscano and McMurray manipulated both the proximal (adjacent) and the distal context (further removed sentential context) simultaneously, making it hard to distinguish between the two effects. Despite these methodological differences, Toscano and McMurray as well as Reinisch and Sjerps both point towards an early time course of distal rate effects.

Recently, two other eye-tracking studies tested the time-course of effects of speech rate, this time on morphosyntactic gender marking in Dutch (Kaufeld, Naumann, Meyer, Bosker, & Martin, 2019; Kaufeld, Ravenschlag, Meyer, Martin, & Bosker, 2019). Just as Reinisch and Sjerps (2013) and Toscano and McMurray (2015), the two studies by Kaufeld et al. found effects of distal speech rate in an early time window after target vowel offset (i.e., 200 ms and 250 ms, respectively; ca. 250–350 ms after vowel onset).

The fact that distal rate effects arise very early in perception has been taken as evidence for the involvement of general auditory mechanisms (Reinisch & Sjerps, 2013; Toscano & McMurray, 2015). More evidence for distal speech rate effects involving general auditory mechanisms comes from findings showing that listeners use the context in perception of a target even when the context is produced by a different talker (Newman & Sawusch, 2009). Even the fast or slow speech rate of one's own voice can change how subsequent other-produced speech sounds are perceived (Bosker, 2017b). Distal speech rate effects have also been found to be insensitive to cognitive load manipulations (Bosker, Reinisch, & Sjerps, 2017), and to be immune to attentional modulation (Bosker, Sjerps, & Reinisch, in press). Moreover, distal rate effects are also induced by non-speech auditory contexts (Bosker, 2017a; Gordon, 1988; Wade & Holt, 2005), by visual cues to speech rate (Bosker, Peeters, & Holler, in press), and are not task-driven, taking place even without explicit attention being drawn to the ambiguous target word (Maslowski, Meyer, & Bosker, 2019b). These findings all argue for distal rate effects to involve mechanisms that arise very early on in speech perception. Note, however, that effects of the distal rate context on word segmentation (e.g., the lexical rate effect) seem to be speech-specific: Pitt, Szostak, and Dilley (2016) found that unintelligible precursor sentences did not induce a lexical rate effect. We return to this issue in the General Discussion.

Different prerequisites have been found for global speech rate effects. Global speech rate tracking is subject to constraints that have not been found for distal speech rate effects. Firstly, global rate tracking is talker-specific, whereas distal rate tracking is talker-independent. Maslowski et al. (2019a) conducted an experiment providing evidence for this. In this experiment, two groups of participants listened to two talkers (A and B) each speaking at two rates (fast and neutral in the high-rate group; slow and neutral in the low-rate group). With considerable rate variation within each talker's speech, no global rate effect was found on perception of the /ɑ, a:/ vowel contrast. The authors interpreted this as the global rate effect being driven by talker-consistent habitual speech rates. That is, global speech rate effects are observed only when talkers show distinct habitual speech rates, but global rate tracking fails with a reasonable amount of intra-talker

variation.

Another difference between the distal and global rate effect is that the global speech rate effect is easily overriden by more local variation. Reinisch (2016b) conducted an experiment in which participants were exposed to speech from two female talkers, one of whom spoke fast and the other slowly. At test, Reinisch observed an effect of habitual speech rate when participants categorized isolated words with temporally ambiguous vowels. That is, target words from the fast talker were more often categorized as long vowel words than target words from the slow talker. However, this global habitual rate effect disappeared in the subsequent experiment, in which the same manipulated vowels were embedded in fast and slow context sentences. Thus, listeners used habitual rate as a cue when no other rate information was available in the first experiment, but this effect was overriden by the fast and slow distal context rates in the second experiment.

A third argument for the global speech rate effect being different from the distal rate effect is that the global rate effect is not induced by one's own voice. Bosker (2017b) showed that one's own distal speech rate can affect perception of an immediately following ambiguous target word spoken by another talker. However, one's own speech rate does not affect perception of another talker's speech rate in larger contexts. Maslowski et al. (2018) recruited two groups of participants, one of which was instructed to speak fast and the other to speak slowly. The two groups were compared on their perception of /ɑ, a:/ words embedded in a neutral Talker A's speech. They found no global rate effects for self-produced global contexts (Experiment 1). Even playback of one's own voice (i.e., passive listening) did not induce a global speech rate effect (Experiment 2). Only when participants listened to speech that was not self-produced (Experiment 3), an effect of global rate was found.

Therefore, global rate effects are constrained by higher-level information such as a talker's habitual rate. Listeners disregard their own speech rate and unreliable habitual rates of others when taking global context into account. Such constraints do not seem to apply to effects of distal speech rate information. Therefore, distal and global speech rate processing have been suggested to involve distinct processing mechanisms. Specifically, Bosker et al. (2017) proposed that speech rate effects take place at two hierarchical stages in a normalization framework of acoustic context effects, such as spectral and rate normalization. The first stage involves early and automatic perceptual auditory processes. Since distal speech rate effects are impervious to talker changes, attentional modulation, and the speech/non-speech nature of the sound context, distal rate effects happen at this early and automatic stage. The second stage involves cognitive adjustments that take place later. These adjustments are conditioned by signal-extrinsic and indexical higher-level information (rather than perceptual normalization of signal-intrinsic cues), such as the identity of the talker (Maslowski et al., 2018; Reinisch, 2016b), the habitual speech rate of the talker (Maslowski et al., 2019a; Reinisch, 2016b), the language that is spoken (Bosker & Reinisch, 2017), the speech register (Reinisch, 2016a), and situation-specific expectations (Bosker et al., 2017). During the second stage, listeners make a categorization decision by comparing the auditory input to a word's expected realization, given contextual factors such as a talker's habitual speech rate (this second stage is comparable to the Computing Cues Relative to Expectations (C-

CuRE) model presented in McMurray & Jongman, 2011, in which the relative importance of speech cues varies with context).

Considering that global rate effects are sensitive to talker identity and stable habitual rates, this entails, according to the two-stage model by Bosker et al. (2017), that global rate effects arise at the second stage, involving cognitive adjustments, while distal rate effects arise at the first stage, involving perceptual normalization. This might also be evident in the time courses of both effects. The two-stage model predicts that distal and global speech rate effects happen in distinct time windows, with global rate influencing perception later in time than distal rate. The time course of the global rate effect has never been assessed directly, nor has it been compared to that of the distal rate effect. Here, we investigated the time courses of both the distal and global rate effects using an eye-tracking paradigm.

The present experiment mimicked Maslowski et al.'s (2019a) categorization experiment on inter-talker variation (i.e., Talker A speaking at one speech rate and Talker B at another). The experimental design and materials were adopted from Maslowski et al. (2019a). The current experiment goes beyond that experiment through the addition of measures of eye fixations, enabling us to investigate the time course of global and distal speech rate effects by analyzing when participants looked at an orthographic target word (cf. Reinisch & Sjerps, 2013). Specifically, a high-rate group listened to neutral speech rate sentences from Talker A and fast sentences from Talker B (i.e., the average rate across talkers was high), while a low-rate group listened to neutral Talker A, but to Talker B speaking at a slow rate (i.e., the average speech rate was low). Their task was to categorize /ɑ, a:/ words embedded in these rate-manipulated sentences (with fixed-rate proximal contexts). During sound presentation, the participants' eye movements and fixations on the two members of a minimal pair were recorded. If global rate effects arise later than distal rate effects, this should become apparent in the participants' eye-tracking data.

Concretely, we predicted that *within groups* the relatively faster rates would induce more long /a:/ responses than the relatively slower rates: In the high-rate group, fast speech should induce more long /a:/ responses than neutral rate speech, and in the low-rate group, neutral rate speech should induce more long /a:/ responses than slow speech. This within-groups distal rate effect should be reflected in more looks to the word with long /a:/ in the relatively faster rates within the two groups. Moreover, based on Reinisch and Sjerps (2013), Toscano and McMurray (2015), Kaufeld, Ravenschlag, et al. (2019), and Kaufeld, Naumann, et al. (2019) we predicted that the distal rate effect should arise very rapidly after vowel offset. Additionally, we predicted that *across groups* a difference in looking patterns would arise in the neutral rate condition: Participants in the low-rate group should show more looks to long targets compared to participants in the high-rate group. This would provide evidence for a global rate effect in the eye-tracking data. Following the two-stage model by Bosker et al. (2017), this global rate effect was predicted to arise only after the distal effect because it involves more higher-level cognitive adjustments.

Method

Participants

42 native Dutch participants (female = 33, M_{age} = 23 years, range = 18–28 years) with normal hearing and vision were recruited from the Max Planck Institute participant pool. All participants gave informed consent to participation. Ethical approval of the study was provided by the Ethics Committee of the Social Sciences faculty of Radboud University (project code: ECSW2014-1003-196). Just as in previous experiments using the same stimuli (Maslowski et al., 2019a, 2018), it was decided a priori to exclude participants for whom the stimuli were insufficiently ambiguous, that is, when they categorized all vowels as being from the same category more than 90% of the time. Eight participants had to be excluded based on this criterion (high-rate group = 5; low-rate group = 3; all eight participants showed > 90% long /a:/ responses). Two other participants were excluded because of technical difficulties. This resulted in two groups of 16 participants each: a high-rate group (female = 12, M_{age} = 23, range = 20–28), who heard fast and neutral speech rates, and a low-rate group (female = 13, M_{age} = 23, range = 20–28), who heard slow and neutral speech rates. With a sample size of 32 participants, we had a power of .95 to observe a global rate effect of > 80% of the size of the global rate effect obtained in Maslowski et al. (2019a) using the same stimuli (see Brehm & Goldrick, 2017, for simulating sample data to estimate power).

Design and materials

The spoken stimuli were taken from Maslowski et al. (2019a). The materials consisted of two minimal pairs differing only in their vowel (*stad/staat*, /stɑt, sta:t/, "city"/"state" and *takje/taakje*, /tɑkjə, ta:kjə/, "twig"/"task"), each embedded in four Dutch context sentences (all containing 24 syllables) without any other instances of /ɑ/ and /a:/. Neither member of a word pair was favoured by the semantic context of the sentence (e.g., *Femke lette goed op of ze niet ging stotteren en toen heeft ze eens "stad/staat" tegen Roos gezegd*, "Femke took care not to stutter and then she said 'city/state' to Roos once"; see Appendix for all sentences and English paraphrases). All sentences were recorded by a native Dutch female and a native Dutch male talker, to increase the salience of talker voice differences. Recordings were divided into target words, buffers (i.e., three syllables before and one syllable after the target word to control for influences of proximal rate; $M_{pre-buffer}$ = 538 ms, $M_{post-buffer}$ = 247 ms), and context sentences (i.e., all speech up to the first buffer and all speech following the second buffer; see formatting in Appendix). Context sentences produced by the two talkers were set to the mean of their durations with PSOLA in Praat (Boersma & Weenink, 2015), such that they were matched in duration across the two talkers. Context sentences were then rate manipulated through linear expansion (factor of 1.6) and compression (factor of 1/1.6 = 0.625) with PSOLA, resulting in three context speech rates: slow ($M_{pre-carrier}$ = 4106 ms, $M_{post-carrier}$ = 1195 ms), neutral (no further rate manipulation; $M_{pre-carrier}$ = 2566 ms, $M_{post-carrier}$ = 747 ms), and fast ($M_{pre-carrier}$ = 1604 ms, $M_{post-carrier}$ = 467 ms).

Spoken target words were excised and manipulated to create two duration continua, ranging from more /ɑ/-like to more /a:/-like perception. It was important to create a continuum that (i) spanned a sufficiently large perceptual range (otherwise the categorization task would be too difficult) as well as (ii) was sufficiently ambiguous to be able to reliably assess context effects in two directions: increasing and decreasing the proportion of long /a:/ responses. The Dutch vowel contrast /ɑ, a:/ is distinguished by both temporal and spectral cues, with /ɑ/ having a shorter duration and a lower F1/F2 than /a:/ (Adank, Van Hout, & Smits, 2004). Therefore, five-step vowel duration continua ranging from 80 to 120 ms (in steps of 10 ms) with ambiguous spectral information (perceptually midway between /ɑ/ and /a:/) were created. First, one long vowel /a:/ was extracted for each talker and durations were manipulated using PSOLA. Then, the F1s and F2s from both talkers were computed and set to fixed ambiguous values with Burg's LPC algorithm as implemented in Praat. The male talker's F1 was 764 Hz and the F2 was 1261 Hz, and the female talker's F1 was 728 Hz and the F2 was 1327 Hz. Finally, the ambiguous vowels were spliced into their consonantal frames /st_t/ and /t_k/. The final set of auditory stimuli was created by concatenating the rate-manipulated context sentences, the original buffer intervals, and the manipulated target words. This resulted in 240 unique stimulus sentences, crossing eight context phrases with three rates, five vowel durations, and two talkers.

The visual targets on the screen were the two members of a minimal pair (e.g., *stad* and *staat*), presented orthographically in Arial, font size 16. In a traditional visual world paradigm, the screen displays different objects or scenes (Allopenna, Magnuson, & Tanenhaus, 1998; Altmann & Kamide, 1999), but eye-tracking paradigms have also been used with orthographic words instead of pictures (McQueen & Viebahn, 2007; Huettig & McQueen, 2007; Huettig, Rommers, & Meyer, 2011). We followed Reinisch and Sjerps (2013) in our use of two referents presented in orthographic form.

Participants were allocated to either the high-rate or the low-rate group. The high-rate group was presented with 40 fast and 40 neutral auditory stimuli (eight sentences × two rates/talkers × five vowel duration targets) with the corresponding visual printed-word stimuli. The 80 auditory items were randomized within each of five blocks. The low-rate group heard 40 slow and 40 neutral items, which were also presented in randomized order in each of five blocks. As such, the high-rate group and the low-rate group listened to the same neutral speech from one talker, but to different rates from the other talker. Which talker (male/female) spoke at a neutral rate was counterbalanced between participants.

Procedure

In the experiment, participants were presented with an auditory stimulus while they looked at an 50.8 cm × 28.6 cm experimental screen with two written targets. The experiment was controlled using Presentation software (v16.5; Neurobehavioral Systems, Albany, CA, USA) combined with a tower-mounted EyeLink 1000 system (SR Research Ltd., Ottawa, Ontario, Canada) sampling at 1000 Hz. Participants were tested individually in a sound-attenuating booth, listening to the auditory stimuli over headphones. Before the start of the experiment, the eye-tracker was adjusted to a height that was comfortable for the participant,

after which the system was calibrated. Eye-tracking data were obtained from participants' right eyes from stimulus onset until stimulus offset plus 1000 ms.

For both groups, the experiment started with instructions, followed by a practice round of eight trials that allowed participants to familiarize themselves with the experimental sentences and speech rates. In the high-rate group, four practice items had a fast speech rate and the other four had a neutral speech rate. In the low-rate group, four practice items had a slow speech rate and four a neutral speech rate. Target words in the practice items contained vowel tokens from the extremes of the duration continua (i.e., 80 and 120 ms) in order to emphasize the vowel contrast.

Each trial started with a fixation cross presented for 300 ms, followed by presentation of two written words as response options in black (either "takje" and "taakje" or "stad" and "staat") at sound onset. The short /ɑ/-word was always shown on one side of the screen and the long /a:/-word was always shown on the other side of the screen. The position of response options (left/right) was counterbalanced between participants. The response options were shown during the whole trial until 1000 ms after sound offset. Participants were instructed to press either "1" on a regular keyboard for the word shown on the left of the screen or "0" for the word on the right side of the screen, thus categorizing the ambiguous target words. The response options were present on the screen from sound onset, but participants were instructed to respond only after they had heard the target word. At button press, the chosen response turned yellow until the end of the trial. If no response was given until 1000 ms after sound offset, a missing response was recorded. The session lasted approximately 50 minutes for the high-rate group and 70 minutes for the low-rate group.

Results

Categorization data

Before analysis, individual participants' categorization responses were inspected visually, to establish that they were not randomly guessing what they had heard. All participants showed a categorization pattern that showed sensitivity to the vowel continua (i.e., a positive-going slope, as shown in Figure 1). The positive slopes of the lines in the figure show that participants more often indicated having heard a long vowel when the absolute durations of vowels were longer. Additionally, within each group, participants responded differently to the same vowels, depending on the distal speech rate context in which they were embedded, as depicted by the different line types. That is, within each group, the relatively faster rate (fast in high-rate group; neutral in low-rate group) induced more long /a:/ responses. Moreover, the figure suggests a difference in the perception of the neutral rate condition between groups, with a higher proportion of long /a:/ responses in the neutral rate for the low-rate group than for the high-rate group, as illustrated by the separation between the two lines in the middle.

[Figure 1 about here.]

We ran logistic Generalized Linear Mixed Models (GLMMs) from the `lme4` package (Bates, Mächler, Bolker, & Walker, 2015) in `R` (R Core Team, 2014) on the categorization data (0.73% missing responses excluded)[1]. Specifically, we performed a two-step statistical analysis. The first step involved a model that matched our experimental design, crossing (within-group) rate conditions with the two participants groups. This first step allowed for assessing statistical evidence for a (within-group) distal rate effect, but it could not test for evidence for a global rate effect. Therefore, a second analysis step involved subsetting the data to only the two neutral rate conditions, comparing performance between the two groups.

The first analysis step involved coding the various rate conditions with respect to the 'relative rate' within each group to measure the distal rate effect. That is, in the high-rate group, fast speech was coded as 'relatively high rate' and neutral speech was coded as 'relatively low rate'. Similarly, in the low-rate group, neutral speech was coded as 'relatively high rate' and slow speech was coded as 'relatively low rate'. As such, the distal rate effect was measured by a comparison of relatively high rates versus relatively low rates. Note that with this coding the neutral speech rate is 'relatively low' in the high-rate group but 'relatively high' in the low-rate group. In addition to the fixed effect Relative Rate (categorical predictor; sum-to-zero coded: relatively high as -0.5, relatively low as 0.5), the model included Group as a categorical predictor (sum-to-zero coded: high-rate group as -0.5, low-rate group as 0.5), Vowel Duration as a continuous predictor (centered around the mean and divided by one standard deviation), and Block as a continuous predictor (centered around the mean and divided by one standard deviation). The interaction between Relative Rate and Group was also included in the model. Random intercepts were included for Participant and Item, with random slopes for Relative Rate, Vowel Duration, and Block by Participant, and for all fixed effects by Item. The random effects structure of the model was arrived at by log-likelihood model comparison of this model to simpler models, with fewer random slopes. Also, including interactions as random slopes resulted in non-convergence errors and were therefore dropped. This resulted in the following model, given here in `R` notation:

$$\begin{aligned} glmer(Vowel\_categorization \sim 1 + Relative\_Rate * Group + Vowel\_Duration + Block + \\ (1 + Relative\_Rate + Vowel\_Duration + Block | Participant) + \\ (1 + Relative\_Rate + Group + Vowel\_Duration + Block | Item)) \end{aligned} \tag{1}$$

The proportion of long /a:/ responses differed significantly with Vowel Duration; long vowel categorization increased for longer durations ($\beta = 1.243, z = 10.358, p < 0.001$). The factor Relative Rate was also significant ($\beta = -2.563, z = -7.572, p < 0.001$), showing that the relatively high rates (fast in the high-rate group; neutral in the low-rate group) received more long /a:/ responses than the relatively low rates (neutral in the high-rate group; slow in the low-rate group). This indicates the presence of a distal rate effect. The interaction between Relative Rate and Group did not reach significance ($\beta = -0.210, z = -0.408, p = 0.683$), demonstrating that this distal rate effect was comparable between groups. No significant main effect of Group

[1] All data have been made available at https://osf.io/c9fyd/.

($\beta = -0.872, z = -1.766, p = 0.077$) was found, although there was a trend for the high-rate group to report hearing more long /a:/ vowels than the low-rate group. Finally, no significant effect of Block was found ($\beta = -0.128, z = -1.124, p = 0.214$), showing that vowel categorization responses did not change over time.

Note, however, that this model (1) cannot inform us about the presence or absence of statistical evidence for any global rate effect. Specifically, neither the main effect of Group nor the interaction between Group and Relative Rate can provide definitive statistical evidence for a global rate effect. That is, the trend towards a main effect of Group (observed above) could, in principle, be driven by an overall difference between the high-rate and low-rate groups, that does not need to rely on the neutral rate conditions per se. Also, any potential interaction between Group and Relative Rate would not provide evidence for a global rate effect, because, again, that could simply be driven by differences in categorization between the fast rate and the slow rate conditions in the two groups – independent from the categorization behavior in neutral rate conditions.

Therefore, for the between-groups global rate effect, we ran another analysis on a subset of the data containing only the neutral rate data. This GLMM included Group as a categorical predictor (sum-to-zero coded: high-rate group as -0.5, low-rate group as 0.5), Vowel Duration, and Block (coded as before), resulting in the model:

$$\begin{aligned} glmer(Vowel\_categorization \sim 1 + Group + Vowel\_Duration + Block + \\ (1 + Vowel\_Duration + Block | Participant) + \\ (1 + Group + Vowel\_Duration + Block | Item)) \end{aligned} \tag{2}$$

Just as the above-mentioned model, the outcomes of this model demonstrated a significant effect of Vowel Duration ($\beta = 1.348, z = 9.040, p < 0.001$). The model also revealed an effect of Group ($\beta = 1.553, z = 2.930, p = 0.003$), with the high-rate group reporting fewer long /a:/ vowel responses in neutral rate speech than the low-rate group (i.e., a global rate effect). There was no significant effect of Block ($\beta = -0.060, z = -0.497, p = 0.619$).

In sum, the results from the button-press categorization responses show a within-groups effect of distal speech rate context; within groups, participants categorized target vowels more often as /a:/ when they were embedded in a relatively faster speech rate as compared to a relatively slower speech rate. Crucially, the results also demonstrate a between-groups global speech rate effect. The low-rate group categorized vowels in neutral rate speech more often as /a:/ than the high-rate group. That is, when neutral rate from one talker sounds relatively fast compared to the speech from another talker, perception of target words in neutral rate is biased towards hearing long vowel target words. Likewise, when a neutral rate seems slow, perception of target words in neutral rate is biased to hearing a shorter target word. This global rate effect in the categorization data replicates the results reported in Maslowski et al. (2019a, 2018). Thus, we proceed with our analysis of the eye-tracking data to assess the time courses of the distal and global rate effects.

Eye fixations

The raw eye-tracking data from each participant were down-sampled from 1000 Hz to 250 Hz for simplicity. Samples with blinks and saccades were excluded from analysis. The areas of interest were set at 300×300 pixels around the center point of each target word. We only analyzed fixations to these interest areas. Hence, fixation proportions were calculated against the fixations to the two interest areas, not to the total number of fixations.

Figure 2 depicts the proportions of fixations to long /a:/ target words for each vowel duration (80–120 ms), collapsing across the rate conditions and groups. This figure shows that the longer the duration of the vowel, the more participants fixated on the target with a long vowel. Figure 3 shows the proportions of long /a:/ target word fixations as a function of the context speech rate in which the target word was embedded (collapsing across vowel durations), roughly reflecting the outcomes as illustrated in Figure 1. Figure 3 suggests that, within groups, participants were more likely to look at the long vowel words if the context speech rate was relatively fast (i.e., fast in the high-rate group; neutral in the low-rate group) than if the speech rate was relatively slow (i.e., neutral in the high-rate group; slow in the low-rate group). Also, it suggests that, across groups, participants' gaze patterns differed in the neutral rate conditions depending on the group: The low-rate group shows more fixations to long /a:/ targets in the neutral rate condition than the high-rate group. Moreover, Figure 3 also suggests a difference between the time courses of distal speech rate effects (relatively high vs. relatively low speech rates) and global speech rate effects (neutral speech rate in high-rate group vs. low-rate group), with the two middle lines diverging at a later time point compared to the within-groups divergences.

[Figure 2 about here.]

[Figure 3 about here.]

Overall analysis

Before the main analysis of the time courses of the distal and global speech rate effects, we determined whether the predicted effects were present in the eye gaze data at all, reflecting the effects in the categorization responses. To statistically test the eye gaze data, we defined a time window of interest starting from 200 ms after target vowel offset. Target vowel offset is the earliest time point at which listeners can access the duration of the target vowel and 200 ms is the time it takes to program and launch a saccade (Altmann & Kamide, 1999). The time window ended at 1250 ms after target vowel offset, since visual inspection showed stabilization of gaze patterns after around 1000 ms.

To test the influences of distal and global speech rates on the looks to long vowel words, the logit-transformed proportions of eye fixations on the long /a:/ target word were quantified with an LMM in `R`. We performed the same two-stage analysis as reported for the behavioral categorization data. The structure of the first model was identical to model (1) used to test the behavioral button-press responses. It included

Relative Rate, Group, Vowel Duration, and Block, as well as the interaction between Relative Rate and Group. The random effects structure included random intercepts for Participant and Item as well as random slope terms for Relative Rate, Vowel Duration, and Block by Participant, and for all four fixed effects by Item, resulting in the model:

$$\begin{aligned} lmer(Logit\_eye\_fixations \sim 1 + Relative\_Rate * Group + Vowel\_Duration + Block + \\ (1 + Relative\_Rate + Vowel\_Duration + Block | Participant) + \\ (1 + Relative\_Rate + Group + Vowel\_Duration + Block | Item)) \end{aligned} \tag{3}$$

The model showed a significant effect of Vowel Duration ($\beta = 0.588, df = 51.6, t = 6.041, p < 0.001$), with more fixations to the long vowel word with increasing vowel durations. The within-groups factor Relative Rate was also significant ($\beta = -1.736, df = 38.5, t = -6.397, p < 0.001$); there were more looks to the long target word in the relatively higher speech rates compared to the lower speech rates, demonstrating an effect of distal speech rate. Group did not significantly affect looks to the long target word ($\beta = -0.365, df = 32.0, t = -0.936, p = 0.357$), nor did Block ($\beta = 0.054, df = 32.2, t = 0.650, p = 0.520$). The interaction between Relative Rate and Group did not reach significance ($\beta = -0.235, df = 32.9, t = -0.526, p = 0.603$), indicating that the distal rate effect did not differ between groups.

Because model (3) did not test for global rate effects, the second stage involved another analysis, run on a subset of the data containing only the neutral rate trials. This second model tested whether global speech rate influenced participants' gaze patterns. The predictors included were identical to those in model (2) testing the global rate effect in the button-press responses. Thus, the model comprised of the fixed effects Group, Vowel Duration, and Block, and the random effects Participant and Item with random slopes for the fixed effects, just like in the corresponding model for the button-press responses:

$$\begin{aligned} lmer(Logit\_eye\_fixations \sim 1 + Group + Vowel\_Duration + Block + \\ (1 + Vowel\_Duration + Block | Participant) + \\ (1 + Group + Vowel\_Duration + Block | Item)) \end{aligned} \tag{4}$$

Vowel Duration had a significant effect on fixations to the long target word ($\beta = 0.791, df = 50.1, t = 6.277, p < 0.001$), in accordance with the model above. Group significantly affected target word fixations ($\beta = 1.021, df = 32.4, t = 2.354, p = 0.025$), indicating an effect of the global context speech rate in the expected direction: The high-rate group showed a lower proportion of looks to the long target word in neutral speech than the low-rate group. Finally, Block had no significant effect on participants' eye fixations ($\beta = 0.084, df = 33.2, t = 0.837, p = 0.409$).

The first of these two models tested within-group differences of speech rate context. It demonstrated fixation differences within groups depending on the within-sentence speech rate (i.e., distal rate effect). Moreover, the second model, testing between-group differences of speech rate context, showed a difference

between the high-rate group and the low-rate group in the fixations to the long target word in the neutral speech conditions (i.e., global rate effect). These results corroborate the distal and global rate effects in the categorization data reported above.

Time course analysis

To statistically test when, in the time window of interest, particular effects arose, is not straightforward. Statistical methods typically used to measure time courses, such as growth curve analyses, cannot compare the onsets of different effects in time series data like the present eye-tracking data. Eye-tracking data are typically a product of sampling from multiple random factors (e.g., participants and items; Baayen, Davidson, & Bates, 2008). Moreover, an auto-correlational structure underlies these densely sampled time series (Seedorff, Oleson, & McMurray, 2018; Cho, Brown-Schmidt, & Lee, 2018) and, hence, adjacent samples in time are generally the product of the same physiological events. At present, there is no single statistical tool that overcomes all these analytical factors (as comprehensively described in Seedorff et al., 2018). We selected the Bootstrapped Differences of Timeseries (BDOTS; Seedorff et al., 2018) approach because it has been designed specifically for the type of data currently under analysis (densely sampled timeseries obtained from eye-tracking). However, we also assessed the time point of the onset of effects by means of the divergence metric implemented in the `R` package `eyetrackingR` (version 0.1.8; Dink & Ferguson, 2015). This divergence analysis is reported in the Supplementary Materials. To summarize, its outcomes were similar to the outcomes of the BDOTS approach reported here.

A complete mathematical treatment of the BDOTS approach is provided in Oleson, Cavanaugh, McMurray, and Brown (2017). We used the BDOTS approach that is implemented in the `R` package `BDOTS` (version 0.1.19; Seedorff et al., 2018). It allows for statistical assessment of the onset of a particular effect as a comparison between two conditions in eye-tracking timeseries data. We tested when in time three effects could be reliably detected. The first effect was the within-groups effect of vowel duration, tested by comparing looks to the long /a:/ word in trials with 80 ms vowels vs. 120 ms vowels. The second effect was the within-groups effect of distal rate. In order to measure the time course of the distal rate effect, the different rate conditions were coded with respect to the 'relative rate' within each group, as explained before. Finally, the third effect was the between-groups effect of global rate, tested by comparing the two neutral rate conditions only (high|neutral vs. low|neutral). Table 1 summarizes the BDOTS analyses of all three effects.

[Table 1 about here.]

For the vowel duration effect, the BDOTS analysis started by fitting a 4-parameter logistic function to individual listeners' patterns ($N = 32$) of fixations to the long /a:/ word in the 80 ms vs. the 120 ms conditions (in the time window from 200 to 1000 ms from vowel offset). This helps to smooth the data, reducing the influence of idiosyncratic patterns of significance on the outcomes. This fitting stage involved visual comparison of fitted curves to observed data and subsequent refitting. In this fitting stage, 4 out

of the 64 logistic curves were dropped due to poor fitting. The remaining 60 curves were fitted assuming autoregressive error (AR1; $n = 51$), or without AR1 if better fits could be obtained without it ($n = 9$). From these fits, standard errors of the mean and confidence intervals were computed using bootstrapping. Based on these values, t-tests were conducted at each time-step (i.e., every 4 ms with the sampling rate of 250 Hz) with a family-wise error adjustment with a modified Bonferroni corrected significance level that takes into account the auto-correlation among t-statistics (Seedorff et al., 2018). Autocorrelation of the t-statistics was 0.9904, the adjusted alpha was calculated to be 0.0029. This analysis demonstrated a single region of significance, starting at 296 ms after vowel offset and ending at 996 ms (i.e., the end of the analyzed time window). This suggests that from 296 ms onwards the looking behavior in trials with 120 ms vowels showed a reliably higher proportion of fixations to the long /a:/ targets than that in trials with 80 ms vowels (see Figure 2).

For the distal rate effect, we fitted 4-parameter logistic functions to individual listeners' patterns of fixations to the long /a:/ word in trials with a 'relatively high rate' (fast in high-rate group; neutral in low-rate group) against a 'relatively low rate' (neutral in high rate group; slow in low rate group) from 200–1000 ms after vowel offset. In the fitting stage, 3 out of the 64 logistic curves were dropped due to poor fitting. Of the remaining 61 curves, 54 were fitted with AR1, and 7 without AR1. In the bootstrapping stage, autocorrelation of the t-statistics was 0.9912, the adjusted alpha was calculated to be 0.004. This analysis detected a single region of significance, starting at 308 ms after vowel offset and ending at 996 ms. That is, from 308 ms onwards, the relatively faster rate showed a reliably higher proportion of fixations to the long /a:/ words than the relatively slower rate (see Figure 3).

For the global rate effect, we fitted 4-parameter logistic functions to the observed fixations to long /a:/ words from 200-1000 ms after vowel offset. However, because the global rate effect concerns a between-participants and between-groups comparison, logistic functions were fit to individual items' patterns of fixations (i.e., a within-items analysis for 40 items: 8 sentences combined with 5 vowel durations), comparing looking behavior in the two neutral rate conditions only (high|neutral vs. low|neutral). All except one curve showed good fitting (66 with AR1; 13 without AR1). In the bootstrapping stage, autocorrelation of the t-statistics was 0.9994, the adjusted alpha was calculated to be 0.0218. This analysis detected a single region of significance, starting at 532 ms after vowel offset and ending at 996 ms. That is, from 532 ms onward, the neutral rate in the low rate group showed a reliably higher proportion of fixations to the long /a:/ words than the neutral rate in the high rate group (see Figure 3).

To summarize, the various BDOTS analyses indicated that (1) increasing the length of the target vowel (i.e., the vowel duration effect) reliably influenced participants' looking behavior from 296 ms after vowel offset onward; (2) hearing the target vowels after a relatively fast speech rate vs. a relatively slow speech rate (i.e., the distal rate effect) likewise reliably influenced participants' looking behavior from a relatively early point in time, namely 308 ms after vowel offset. However, (3) hearing the target vowels in the low-rate group vs. the high-rate group (i.e., the global rate effect) only reliably influenced participants' looking behavior

from 532 ms onward.

Jackknife analysis

Although the time course analysis above demonstrated at which point in time different effects could reliably be detected, it does not inform us about whether one effect arose significantly earlier or later in time than another effect. Therefore, we performed a jackknife analysis, focusing on the contrast between the distal effect and the effect of vowel duration serving as a baseline, and on the contrast between the distal and global effect as our primary contrast of interest. This jackknife analysis was based on the time points at which the various effects reached certain percentages of their maxima, thus allowing comparison across different types of effects, even when they differ in their absolute size (following Reinisch & Sjerps, 2013; Toscano & McMurray, 2015).

Fixation proportions on the long /a:/ words for every time sample (i.e., in 4 ms bins) were logit transformed. The within-groups vowel duration effect was quantified following Toscano & McMurray, 2015, by computing linear regressions between the (logit transformed) fixation proportions on the long /a:/ words and vowel duration step. The slope was used as a measure of the size of the vowel duration effect. To quantify the within-groups distal rate effect, the different rate conditions were coded with respect to the 'relative rate' within each group (i.e., the same coding used in the BDOTS time course analysis above). As such, the distal rate effect was measured by subtracting the transformed proportions of the 'relatively low rate' condition (neutral in high rate group; slow in low rate group) from those of the 'relatively high rate' condition (fast in high-rate group; neutral in low-rate group). Finally, to quantify the between-groups global rate effect, we subtracted the transformed proportions of the high|neutral condition from those of the low|neutral condition (i.e., involving only the two neutral rate conditions). All effect size measures were then smoothed by applying an 80 ms asymmetrical sawtooth window (as did McMurray, Clayards, Tanenhaus, & Aslin, 2008; Reinisch & Sjerps, 2013), weighing a given sample's value for the range of values from 80 ms before that sample (with more distant values contributing less). Finally, the smoothed measures were divided by the maximum value in order to normalize the timecourses of the effects.

In order to statistically compare the time points at which differents effects arose, we performed a jackknife procedure (Ulrich & Miller, 2001). After the normalization described above, we computed when in time the normalized effects crossed specific thresholds. That is, we calculated time points at every 10% step of the maxima until 80% of the maxima was reached (cf. Reinisch & Sjerps, 2013) within a time window 200–1000 ms after vowel offset. This was computed for multiple subsets of the total number of items (i.e., involving a within-items analysis, considering that the global rate effect concerned a between-subjects comparison). Each subset contained data from N-1 items; that is, each item was excluded once. Two ANOVAs (one for the distal vs. vowel duration contrast; one for the distal vs. global contrast) tested the jackknife-datasets of time points at which the different effects reached certain percentages of their maxima, with Effect Type as predictor. Because each item contributed multiple times to the various jackknife-datasets, F-values (and

respective p-values) were adjusted by dividing by $(N-1)^2$ (Reinisch & Sjerps, 2013; Ulrich & Miller, 2001). The p-values were calculated with the same degrees of freedom as for the respective simple dataset (Reinisch & Sjerps, 2013; Ulrich & Miller, 2001).

Figure 4 shows how the effects of vowel duration, distal rate, and global rate increase up to their maxima. The effect of vowel duration and the distal effect seem to arise the earliest. The global effect seems to arise at a later time point. Table 2 reports the time points at which certain percentages of the maxima were reached, together with the adjusted F- and p-values. The statistical comparison of the vowel duration effect and the distal effect did not reveal significant differences, except at the 80% time point (on average, at 652 and 756 ms respectively). Since the 80% time point is a relatively late time point, this suggests that the effects of vowel duration and distal rate both arise at equally early time points (in line with Reinisch & Sjerps, 2013; Toscano & McMurray, 2015; Kaufeld, Ravenschlag, et al., 2019; Kaufeld, Naumann, et al., 2019). However, a significant difference was observed between the time point at which the distal rate effect and the global rate effect reached 20% of their maxima (at 408 and 512 ms, respectively), and a trend at 30% of their maxima (at 481 and 550 ms, respectively). This indicates that the distal rate effect and the global rate effect reached 20% of their maxima at significantly different points in time. That is, the distal rate effect appears at an earlier time compared to the global rate effect.

[Figure 4 about here.]

[Table 2 about here.]

General Discussion

This study aimed to determine and compare the time courses of the distal speech rate effect (effect of the sentence context speech rate) and the global speech rate effect (effect of the speech rate context beyond the sentence) on /ɑ, a:/ perception in Dutch. The experiment tested two groups of participants. One group listened to a neutral rate Talker A and a fast Talker B (i.e., the high-rate group). The other group listened to neutral Talker A, but to Talker B speaking at a slow speech rate (i.e., the low-rate group). Participants performed a two-alternative forced choice task, in which they had to indicate whether they had heard a word with an /ɑ/ or with an /a:/ (e.g., *stad/staat*, /stɑt, sta:t/, "city"/"state"). Additionally, their eye fixations were measured to investigate when they looked at a given written target word on the screen. The distal rate effect was measured by comparing categorization responses and eye fixations within groups (fast vs. neutral in high-rate group; neutral vs. slow in low-rate group), whereas the global speech rate effect was measured by comparing the two between-groups neutral rate conditions.

Regarding the categorization results, we observed a within-group distal speech rate effect in each of the two groups, with relatively faster speech rates (i.e., fast in high-rate group; neutral in low-rate group) receiving more long /a:/ responses compared to relatively slower speech rates (i.e., neutral in high-rate group; slow in low-rate group). Moreover, we found a between-groups effect of global speech rate, with neutral rate

in the low-rate group receiving more long /a:/ responses than neutral rate speech in the high-rate group. Even though these two effects did not demonstrate complete shifts of target perception (i.e., the high|neutral condition always being perceived as 0% /a:/), the effect sizes are in line with previous literature, replicating earlier work on distal and global speech rate effects (Maslowski et al., 2018, 2019a; Reinisch & Sjerps, 2013).

With regard to the time courses of the distal and global speech rate effects, we hypothesized that the effects would arise at different times, rather than manifesting themselves simultaneously. Specifically, we expected the global speech rate effect to emerge later than the distal rate effect. This hypothesis was based on predictions made by Bosker et al.'s (2017) two-stage hierarchical model for acoustic context effects. This model states that acoustic context effects take place at two different stages. The first stage involves perceptual processing of auditory input, which is domain-general, automatic, and obligatory, whereas the second stage takes into account higher-level factors such as talker identity. Because distal speech rate directly affects perceptual processing of temporally ambiguous sounds (e.g., Reinisch et al., 2011), distal rate information is argued to be used at the first stage. The global speech rate effect, however, has been suggested arise at the subsequent stage (Bosker & Ghitza, 2018; Maslowski et al., 2018, 2019a), given that global rate tracking is sensitive to talker-identity (Maslowski et al., 2018) and can be overriden by local speech rate variation (Reinisch, 2016b). The two-stage model therefore predicts that the global rate effect should be observed in a later time window than the distal rate effect.

The results from the eye fixations were consistent with the predictions from the two-stage model; the global rate effect arose considerably later (earliest significant time point was 532 ms after vowel offset) than the distal rate effect (earliest significant time point was 308 ms). Moreover, the jackknife analysis showed that the trajectory of the differences in time course between these two types of rate effects was reliably different at the earliest points in time (i.e., at 20 and 30% of the effects' maxima). In contrast, the distal rate effect arose approximately as early as the effect of manipulating the vowel duration itself. This suggests that distal rate effects arise at an early perceptual stage, while global rate effects arise at a later stage compared to distal rate effects, even when accounting for differences in absolute effect size in the jackknife analysis.

The time course of the distal rate effect is comparable to the time courses of the speech rate effects found by Reinisch and Sjerps (2013), Toscano and McMurray (2015), Kaufeld, Ravenschlag, et al. (2019), and Kaufeld, Naumann, et al. (2019). Kaufeld, Ravenschlag, et al. found effects of distal speech rate after 250 ms after vowel offset. Reinisch and Sjerps and Toscano and McMurray estimated their speech rate effects to start between 300 and 400 ms after target vowel onset (Reinisch and Sjerps) and target word onset respectively (Toscano and McMurray). Given that time point 0 in the current study was target vowel offset, our time line should be shifted approximately 100 ms (i.e., the average vowel duration in the current study) for an accurate comparison. That is, our distal rate effect arose about 400 ms after vowel onset, whereas the global effect arose approximately 630 ms after vowel onset. The timing of our distal rate effect is very similar to the effects reported in Reinisch and Sjerps, Toscano and McMurray, and Kaufeld, Ravenschlag, et al.. However, since the estimated starting points are dependent on, for instance, the width of the time bins chosen, these

specific time points should be considered approximations, rather than precise estimates on a millisecond timescale. Furthermore, note that Toscano and McMurray manipulated proximal (immediately adjacent) and distal (non-adjacent) contexts simultaneously, entailing that their effect could be proximal, distal, or a combination of the two.

Before turning to the implications of our findings, one limitation of our experiment is that we used a Visual World Paradigm with a two-alternative forced choice (2AFC) task with orthographic targets, rather than the standard four-alternative forced choice (4AFC) task using pictures. There are some concerns about the 2AFC paradigm with orthographic targets, such as (a) the issue of visual similarity between targets and (b) the presence of only two referents potentially leading to more categorical responding. However, we argue that our findings are unlikely to be a result of these factors. Our use of orthographic targets that looked very similar may have heightened competition effects between these targets. Yet, because the same displays were used in all conditions, this competition was equal for both types of speech rate effects that we investigated. This allowed us to compare the time courses of the two effects despite the similarity in visual targets.

Regarding the use of a 2AFC task, McMurray, Aslin, Tanenhaus, Spivey, and Subik (2008) reported that the presence of only two objects on a screen may heighten attention to these stimuli in an artificial way. McMurray et al. investigated to which extent listeners are sensitive to within-category acoustic variation by comparing the outcomes of different 2AFC and 4AFC identification tasks. They found shallower identification slopes in their 4AFC tasks than in their 2AFC tasks, particularly in their 2AFC task using non-words. However, their 2AFC task with lexical items demonstrated more sensitivity to within-category detail than the task using non-words, as such providing a more reliable measure for these types of effects. Moreover, our experiment shows that within-category sensitivity can be demonstrated with a 2AFC task, given the fact that fixations to long /a:/ words in our experiment depended on the length of the vowel, with shallow identification slopes, comparable to those in McMurray et al.'s 4AFC tasks. Therefore, we are confident that the present 2AFC task provided an appropriate method for testing our research questions, in line with earlier work (Reinisch & Sjerps, 2013; Kaufeld, Ravenschlag, et al., 2019; Kaufeld, Naumann, et al., 2019; Toscano & McMurray, 2015). Future work may investigate whether other designs lead to similar results.

The current study established the time courses and separability of the distal and the global speech rate effects. Accounts of neural entrainment to speech have attempted to explain distal speech rate effects as a result of neural oscillations in the theta range phase-locking to the (slow and fast) amplitude modulations in the context (Ghitza, 2012; Giraud & Poeppel, 2012; Peelle & Davis, 2012). For instance, Kösem et al. (2018) tested whether neural oscillations can directly shape perception of the following speech signal, using magnetoencephalography (MEG). In their study, participants listened to fast and slow sentences (i.e., distal rate manipulation), followed by a Dutch ambiguous /ɑ-a:/ target word. They found that the brain tracked the speech rhythm of the fast and slow context sentences. Additionally, these fast and slow neural rhythms were observed to persist even after the context sentence had ceased: In the same target time window, evidence for a fast neural rhythm was present when preceded by a fast context sentence, but a slow neural rhythm was

present when preceded by a slow context sentence. Hence, the speech-brain entrainment induced by the fast and slow context sentences carried on for a number of cycles after the rhythm it was entrained to changed. Moreover, the extent to which individuals showed this sustained neural entrainment in the target window was predictive of the behavioral rate effect: Individuals who showed stronger entrainment to the speech rhythm of the distal context also showed a larger perceptual bias in word categorization in the expected direction (i.e., slow-rate entrainment to short-vowel target words and fast-rate entrainment to long-vowel target words). This is in line with psychoacoustic findings that only rhythms in the theta range (3–9 Hz) induce these distal rate effects (Bosker & Ghitza, 2018), and that destroying the rhythm in the distal context also eliminates distal rate effects (Bosker, 2017a). This is accounted for by a temporal sampling framework, whereby entrained theta oscillations impose periodic phases of neuronal excitation and inhibition, thus sampling the input signal at the appropriate temporal granularity.

An important question for further work is whether neural entrainment can also explain global speech rate effects. Previous studies on global speech rate tracking have suggested that their effects had neural correlates similar to the ones found for distal speech rate (Baese-Berk et al., 2014). However, Alexandrou, Saarinen, Kujala, and Salmelin (2018), using MEG, observed spatial differentiation between the neural regions involved in processing global speech rate (temporal cortex bilaterally and right parietal cortex) and those processing distal speech rate variation (left parietal regions). The present outcomes extend the finding of spatially distinct neural regions that underlie global and distal speech rate processing by providing evidence for temporal differentiation as well: We show that global rate effects have a distinct time course in perceptual processing relative to distal rate effects (i.e., arise later). Hence, it would seem unlikely that one and the same neurobiological mechanism could account for both global and distal speech rate processing without additional principles.

In fact, previous research on the global speech rate effect has shown that this global effect is subject to constraints that have not been found for the distal rate effect. Specifically, the global speech rate effect is talker-specific (Maslowski et al., 2019a, 2018; Reinisch, 2016b) and global rate tracking fails with considerable speech rate variation within a given talker (Maslowski et al., 2019a; Reinisch, 2016b), whereas the distal rate effect seems automatic and obligatory (but see Pitt et al., 2016. Thus, while general-auditory mechanisms like sustained neural entrainment have been proposed to underlie the distal rate effect (Bosker, 2017a; Wade & Holt, 2005), the global speech rate effect is unlikely to be explained by domain-general auditory principles.

The findings of the current study corroborate this; the difference in time courses between the global and distal speech rate effects shows that participants took longer to take global speech rate into account, compared to distal speech rate. This indicates that higher-level factors are considered after the first perceptual normalization for distal rate. Consequently, the global rate effect in the present study does not fit in a straightforward manner with current theories of neural entrainment to speech rate, in which brain oscillations adapt to the rhythm of an auditory signal independent from talker identity (Bosker, 2017a).

Therefore, in our view, accounts of neural entrainment (in their present form) are unlikely to suffice

to explain how the talker-specific and relatively late global rate effect influences perception of temporally ambiguous cues. These theories may be adapted to explain the global rate effect by including a system for talker recognition that feeds into the mechanism that tracks the speech envelope. The question remains, however, whether it is plausible that such mechanisms, one for rate tracking and one for talker tracking, feed into each other, since proximal and distal speech rate effects seem to involve general-auditory mechanisms without the need for a system that incorporates feedback about the talker's identity (Newman & Sawusch, 2009; Bosker, 2017b). It may also be the case that, although different types of speech rate effects (e.g., distal vs. global) have typically been described as falling into the same category of rate normalization, they involve distinct computations that are not yet well described.

In fact, the same may be said for how rate-dependent speech processing differentially affects the perception of segmental distinctions (e.g., /ɑ/ vs. /a:/) or lexical perception (e.g., 'disappearing' function words in the lexical rate effect). Considering distal rate effects, the present paper on phonetic boundary shifts, together with a body of literature on rate-dependent segmental perception, suggests that distal rate effects arise very early in time, are very robust, possibly involving automatic domain-general mechanisms. However, data from distal rate effects on lexical perception suggest that the lexical rate effect operates only on intelligible speech (Pitt et al., 2016). Moreover, the one study testing the time course of the lexical rate effect using eye-tracking seems to indicate a relatively later time point at which the lexical rate effect arises (at least 500 ms after critical word offset; Brown et al., 2012). Considering global rate effects, earlier work on phonetic boundary shifts has demonstrated that global rate effects are constrained by signal-extrinsic higher-level factors such as talker identity (Maslowski et al., 2018; Reinisch, 2016b) and within-talker rate variation (Maslowski et al., 2019a). However, the lexical rate effect was observed with an experimental design involving multiple talkers, each talking with variable speech rates (Baese-Berk et al., 2014). What underlies the different behaviors of these two types of rate-dependent effects remains unclear. Bosker (2017a) has speculated that the two phenomena may behave differently, because they operate on different processing levels, with segmental rate effects taking place on a sublexical and domain-general processing level and the lexical rate effect taking place at a lexical domain-specific processing level. This calls for future empirical investigations of the different types of distal and global rate effects, particularly with respect to their time course.

The findings of the current study have implications that go beyond speech rate processing. A large number of studies that targeted the time course of the uptake of speech cues have found an immediate integration of cues, whether they were lower-level perceptual cues (e.g., McMurray, Clayards, et al., 2008; Toscano & McMurray, 2012, 2015; Reinisch & Sjerps, 2013) or higher-level linguistic cues (see e.g., Kingston, Levy, Rysling, & Staub, 2016 on the Ganong effect, Salverda, Dahan, & McQueen, 2003 on prosodic cue integration, and Kaufeld, Ravenschlag, et al., 2019 on morphosyntactic cue integration). However, recently, several studies have found evidence for certain acoustic cues being buffered in memory, suggesting that some acoustic cues may be processed at a later stage in word recognition. For instance, Galle, Klein-Packard, Schreiber, and McMurray (2019) found delays in the use of fricative place of articulation. Similarly, Mitterer,

Kim, and Cho (2019) found the uptake of disambiguating prosodic cues to phoneme identification to take place at a later point in time. Together with the current study, these findings imply that lexical access transpires over multiple stages. These studies provide a fruitful working ground for future work disentangling how and why some speech cues are processed earlier than others.

In sum, this study measured online language processing with an eye-tracking paradigm to test the time course of lexical activation, varying distal and global contextual speech rates. The results illuminate timing differences between the two speech rate effects, supporting the idea that the distal speech rate effect may underlie an early perceptual mechanism, whereas the global speech rate effect may be controlled by a different cognitive adjustment mechanism. This is in line with predictions from the two-stage model of acoustic context effects (Bosker et al., 2017). Future work may investigate which neurobiological mechanisms underlie global speech rate processing, joining the distal and global rate effects that have parallel consequences for speech perception in a single theoretical, neurobiologically plausible framework.

Acknowledgements

We thank Annelies van Wijngaarden for lending her voice. We also thank Maarten van den Heuvel for his technical support on the project. We are grateful to Phillip Alday, Laurel Brehm, Eva Reinisch, and Matthias Sjerps for their statistical advice.

References

Adank, P., Van Hout, R., & Smits, R. (2004). An acoustic description of the vowels of Northern and Southern Standard Dutch. *The Journal of the Acoustical Society of America*, *116*(3), 1729–1738. doi: 10.1121/1.1779271

Alexandrou, A. M., Saarinen, T., Kujala, J., & Salmelin, R. (2018). Cortical tracking of global and local variations of speech rhythm during connected natural speech perception. *Journal of Cognitive Neuroscience*, *30*(11), 1704–1719. doi: 10.1162/jocn_a _01295

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. *Journal of Memory and Language*, *38*(4), 419–439. doi: 10.1006/jmla.1997.2558

Altmann, G. T., & Kamide, Y. (1999). Incremental interpretation at verbs: Restricting the domain of subsequent reference. *Cognition*, *73*(3), 247–264. doi: 10.1016/S0010-0277(99)00059-1

Baayen, R. H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effects modeling with crossed random effects for subjects and items. *Journal of Memory and Language*, *59*(4), 390–412. doi: 10.1016/j.jml.2007.12.005

Baese-Berk, M. M., Dilley, L. C., Henry, M. J., Vinke, L., & Banzina, E. (2019). Not just a function of function words: Distal speech rate influences perception of prosodically weak syllables. *Attention, Perception, & Psychophysics*, *81*(2), 571–589. doi: 10.3758/s13414-018-1626-4

Baese-Berk, M. M., Heffner, C. C., Dilley, L. C., Pitt, M. A., Morrill, T. H., & McAuley, J. D. (2014). Long-term temporal tracking of speech rate affects spoken-word recognition. *Psychological Science*, *25*(8), 1546–1553. doi: 10.1177/0956797614533705

Bates, D., Mächler, M., Bolker, B., & Walker, S. (2015). Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, *67*(1), 1–48. doi: 10.18637/jss.v067.i01

Boersma, P., & Weenink, D. (2015). *Praat: doing phonetics by computer computer program. Version 5.4. 09.* Retrieved from `http://www.praat.org/`

Bosker, H. R. (2017a). Accounting for rate-dependent category boundary shifts in speech perception. *Attention, Perception, & Psychophysics*, *79*(1), 333–343. doi: 10.3758/s13414-016-1206-4

Bosker, H. R. (2017b). How our own speech rate influences our perception of others. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*, 1225–1238. doi: 10.1037/xlm0000381

Bosker, H. R., & Ghitza, O. (2018). Entrained theta oscillations guide perception of subsequent speech: Behavioural evidence from rate normalisation. *Language, Cognition and Neuroscience*, *33*(8), 955–967. doi: 10.1080/23273798.2018.1439179

Bosker, H. R., Peeters, D., & Holler, J. (in press). How visual cues to speech rate influence speech perception. *Quarterly Journal of Experimental Psychology*. (Advance online publication)

Bosker, H. R., & Reinisch, E. (2017). Foreign languages sound fast: Evidence from implicit rate normalization. *Frontiers in Psychology*, *8*, 1063. doi: 10.3389/fpsyg.2017.01063

Bosker, H. R., Reinisch, E., & Sjerps, M. J. (2017). Cognitive load makes speech sound fast, but does not modulate acoustic context effects. *Journal of Memory and Language*, *94*, 166–176. doi: 10.1016/j.jml.2016.12.002

Bosker, H. R., Sjerps, M. J., & Reinisch, E. (in press). Temporal contrast effects in human speech perception are immune to selective attention. *Scientific Reports*. (Advance online publication)

Brehm, L., & Goldrick, M. (2017). Distinguishing discrete and gradient category structure in language: Insights from verb-particle constructions. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *43*(10), 1537–1556. doi: 10.1037/xlm0000390

Brown, M., Dilley, L. C., & Tanenhaus, M. K. (2012). Real-time expectations based on context speech rate can cause words to appear or disappear. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 34).

Cho, S.-J., Brown-Schmidt, S., & Lee, W.-y. (2018). Autoregressive generalized linear mixed effect models with crossed random effects: An application to intensive binary time series eye-tracking data. *Psychometrika*, *83*(3), 751–771. doi: 10.1007/s11336-018-9604-2

Diehl, R. L., & Walsh, M. A. (1989). An auditory basis for the stimulus-length effect in the perception of stops and glides. *The Journal of the Acoustical Society of America*, *85*(5), 2154–2164. doi: 10.1121/1.397864

Dilley, L. C., & Pitt, M. A. (2010). Altering context speech rate can cause words to appear or disappear. *Psychological Science*, *21*(11), 1664–1670. doi: 10.1177/0956797610384743

Dink, J. W., & Ferguson, B. (2015). *eyetrackingR: An R library for eye-tracking data analysis.* Retrieved from `http://www.eyetrackingr.com/`

Galle, M. E., Klein-Packard, J., Schreiber, K., & McMurray, B. (2019). What are you waiting for? real-time integration of cues for fricatives suggests encapsulated auditory memory. *Cognitive Science*, *43*(1), e12700. doi: 10.1111/cogs.12700

Ghitza, O. (2012). On the role of theta-driven syllabic parsing in decoding speech: Intelligibility of speech with a manipulated modulation spectrum. *Frontiers in Psychology*, *3*, 238. doi: 10.3389/fpsyg.2012.00238

Giraud, A.-L., & Poeppel, D. (2012). Cortical oscillations and speech processing: Emerging computational principles and operations. *Nature Neuroscience*, *15*(4), 511–517. doi: 10.1038/nn.3063

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics*, *43*(2), 137–146. doi: 10.3758/BF03214191

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*, *57*(4), 460–482. doi: 10.1016/j.jml.2007.02.001

Huettig, F., Rommers, J., & Meyer, A. S. (2011). Using the visual world paradigm to study language processing: A review and critical evaluation. *Acta Psychologica*, *137*(2), 151–171. doi: 10.1016/j.actpsy.2010.11.003

Kaufeld, G., Naumann, W., Meyer, A. S., Bosker, H. R., & Martin, A. E. (2019). Contextual speech rate influences morphosyntactic prediction and integration. *Language, Cognition and Neuroscience*, 1–16. (Advance online publication) doi: 10.1080/23273798.2019.1701691

Kaufeld, G., Ravenschlag, A., Meyer, A. S., Martin, A. E., & Bosker, H. R. (2019). Knowledge-based and signal-based cues are weighted flexibly during spoken language comprehension. *Journal of Experimental Psychology: Learning, Memory, and Cognition*. (Advance online publication) doi: 10.1037/xlm0000744

Kingston, J., Levy, J., Rysling, A., & Staub, A. (2016). Eye movement evidence for an immediate ganong effect. *Journal of Experimental Psychology: Human Perception and Performance*, *42*(12), 1969–1988.

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. doi: 10.1016/j.cub.2018.07.023

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2018). Listening to yourself is special: Evidence from global speech rate tracking. *PloS one*, *13*(9), e0203571. doi: 10.1371/journal.pone.0203571

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019a). How the tracking of habitual rate influences speech perception. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *45*(1), 128–138. doi: 10.1037/xlm0000579

Maslowski, M., Meyer, A. S., & Bosker, H. R. (2019b). Listeners normalize speech for contextual speech rate even without an explicit recognition task. *The Journal of the Acoustical Society of America*, *146*(1), 179–188. doi: 10.1121/1.5116004

McMurray, B., Aslin, R. N., Tanenhaus, M. K., Spivey, M. J., & Subik, D. (2008). Gradient sensitivity to within-category variation in words and syllables. *Journal of Experimental Psychology: Human Perception and Performance*, *34*(6), 1609–1631. doi: 10.1037/a0011747

McMurray, B., Clayards, M. A., Tanenhaus, M. K., & Aslin, R. N. (2008). Tracking the time course of phonetic cue integration during spoken word recognition. *Psychonomic Bulletin & Review*, *15*(6), 1064–1071. doi: 10.3758/PBR.15.6.1064

McMurray, B., & Jongman, A. (2011). What information is necessary for speech categorization? harnessing variability in the speech signal by integrating cues computed relative to expectations. *Psychological review*, *118*(2), 219–246. doi: 10.1037/a0022325

McQueen, J. M., & Viebahn, M. C. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology*, *60*(5), 661–671. doi: 10.1080/17470210601183890

Miller, J. L., & Baer, T. (1983). Some effects of speaking rate on the production of /b/ and /w/. *The Journal of the Acoustical Society of America*, *73*(5), 1751–1755. doi: 10.1121/1.389399

Mitterer, H., Kim, S., & Cho, T. (2019). The glottal stop between segmental and suprasegmental processing: The case of maltese. *Journal of Memory and Language*, *108*, 104034. doi: 10.1016/j.jml.2019.104034

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics*, *58*(4), 540–560. doi: 10.3758/BF03213089

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics*, *37*(1), 46–65. doi: 10.1016/j.wocn.2008.09.001

Oleson, J. J., Cavanaugh, J. E., McMurray, B., & Brown, G. (2017). Detecting time-specific differences between temporal nonlinear curves: Analyzing data from the visual world paradigm. *Statistical Methods in Medical Research*, *26*(6), 2708–2725. doi: 10.1177/0962280215607411

Peelle, J. E., & Davis, M. H. (2012). Neural oscillations carry speech rhythm through to comprehension. *Frontiers in Psychology*, *3*, 320. doi: 10.3389/fpsyg.2012.00320

Pitt, M. A., Szostak, C., & Dilley, L. C. (2016). Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate. *Attention, Perception, & Psychophysics*, *78*(1), 334–345. doi: 10.3758/s13414-015-0981-7

R Core Team. (2014). R: A language and environment for statistical computing [Computer software manual]. Vienna, Austria. Retrieved from `http://www.R-project.org/`

Reinisch, E. (2016a). Natural fast speech is perceived as faster than linearly time-compressed speech. *Attention, Perception, & Psychophysics*, *78*(4), 1203–1217. doi: 10.3758/s13414-016-1067-x

Reinisch, E. (2016b). Speaker-specific processing and local context information: The case of speaking rate. *Applied Psycholinguistics*, *37*(6), 1397–1415. doi: 10.1017/S0142716415000612

Reinisch, E., Jesse, A., & McQueen, J. M. (2011). Speaking rate from proximal and distal contexts is used

during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *37*(3), 978–996. doi: 10.1037/a0021923

Reinisch, E., & Sjerps, M. J. (2013). The uptake of spectral and temporal cues in vowel perception is rapidly influenced by context. *Journal of Phonetics*, *41*(2), 101–116. doi: 10.1016/j.wocn.2013.01.002

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*(1), 51–89. doi: 10.1016/S0010-0277(03)00139-2

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics*, *62*(2), 285–300. doi: 10.3758/BF03205549

Seedorff, M., Oleson, J., & McMurray, B. (2018). Detecting when timeseries differ: Using the Bootstrapped Differences of Timeseries (BDOTS) to analyze Visual World Paradigm data (and more). *Journal of Memory and Language*, *102*, 55–67. doi: 10.1016/j.jml.2018.05.004

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(5), 1074–1095. doi: 10.1037/0096-1523.7.5.1074

Toscano, J. C., & McMurray, B. (2012). Cue-integration and context effects in speech: Evidence against speaking-rate normalization. *Attention, Perception, & Psychophysics*, *74*(6), 1284–1301. doi: 10.3758/s13414-012-0306-z

Toscano, J. C., & McMurray, B. (2015). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition, and Neuroscience*, *30*(5), 529–543. doi: 10.1080/23273798.2014.946427

Ulrich, R., & Miller, J. (2001). Using the jackknife-based scoring method for measuring lrp onset effects in factorial designs. *Psychophysiology*, *38*(5), 816–827. doi: 10.1111/1469-8986.3850816

Wade, T., & Holt, L. L. (2005). Perceptual effects of preceding nonspeech rate on temporal properties of speech categories. *Perception & Psychophysics*, *67*(6), 939–950. doi: 10.3758/BF03193621

Appendix

Two talkers were recorded producing a set of eight Dutch stimulus sentences (English paraphrase below). These sentences were composed of an /ɑ, a:/ target word, with buffers on either side of the target, and rate-manipulated context phrases (ratio 1.6 for slow, 1 for neutral, and 0.625 for fast). The formatting denotes [context phrase] **buffer** target **buffer** [context phrase].

	Sentences and translations
1	[Peter fluisterde Ilse iets verkeerd in en toen hoorde] **Ilse het** tak-/taak**je** [gezegd worden]. *Peter whispered something in Ilse's ear incorrectly and then Ilse heard "the twig/task" being said.*
2	[Toen Luuk mompelend iets tegen Lotte vertelde hoorde] **Lotte het** tak-/taak**je** [gezegd worden]. *When Luuk muttered something to Lotte, Lotte heard "the twig/task" being said.*
3	[Riet probeerde de notitie te ontcijferen en plots] **kon ze het** tak-/taak**je** [onderscheiden]. *Riet was trying to decipher the note and suddenly she could discern the twig/task.*
4	[Loes twijfelde over de juiste oplossing en toch streep]**te ze het** tak-/taak**je** [door op de toets]. *Loes was unsure about the correct solution and yet she crossed out the twig/task on the test.*
5	[Toen Evelien gisteren iets onnozels wilde zeggen] **heeft ze eens** stad/staat **ge**[zegd tegen Job]. *When Evelien wanted to say something silly yesterday, she said "city/state" to Job once.*
6	[Terwijl Niels rustig zijn tijdschrift stond te lezen hebben de] **heren eens** stad/staat **te**[gen hem gebruld]. *While Niels was peacefully reading his magazine, the gentlemen roared "city/state" to him once.*
7	[Femke lette goed op of ze niet ging stotteren en toen] **heeft ze eens** stad/staat **te**[gen Roos gezegd]. *Femke took care not to stutter and then she said "city/state" to Roos once.*
8	[Toen Simon de oplossing even niet meer wist fluisterde] **Nienke eens** stad/staat **in** [zijn linkeroor]. *Just as Simon could no longer remember the solution, Nienke whispered "city/state" once in his left ear.*

List of Figures

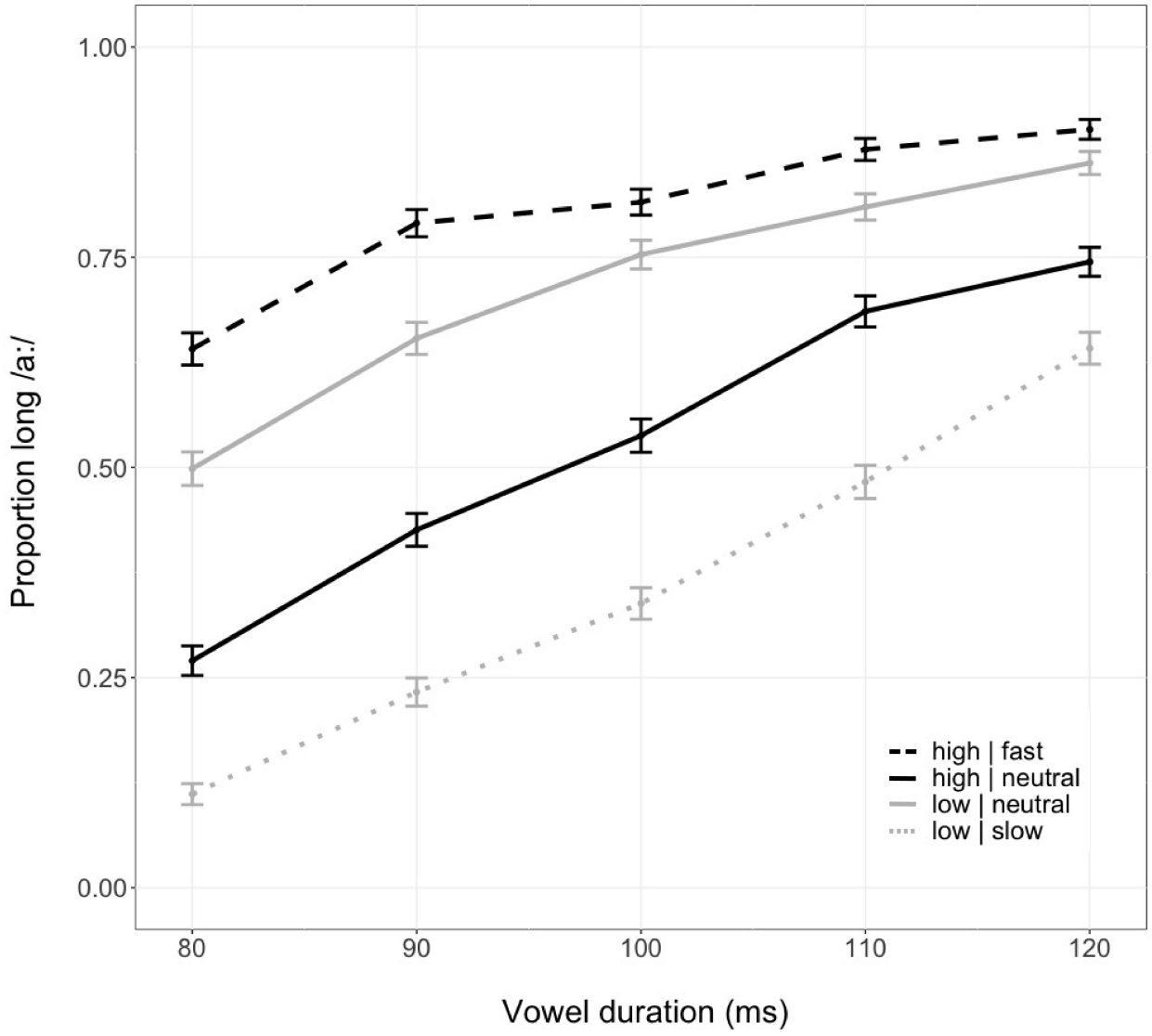

Figure 1: Average categorization data in proportion of long /a:/ responses. The X-axis indicates Vowel Duration (80–120 ms). Color indicates Group, with black representing the high-rate group and gray the low-rate group. Line type indicates Rate Condition, with a dashed line for fast contexts, solid for neutral contexts, and dotted for slow contexts. 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.

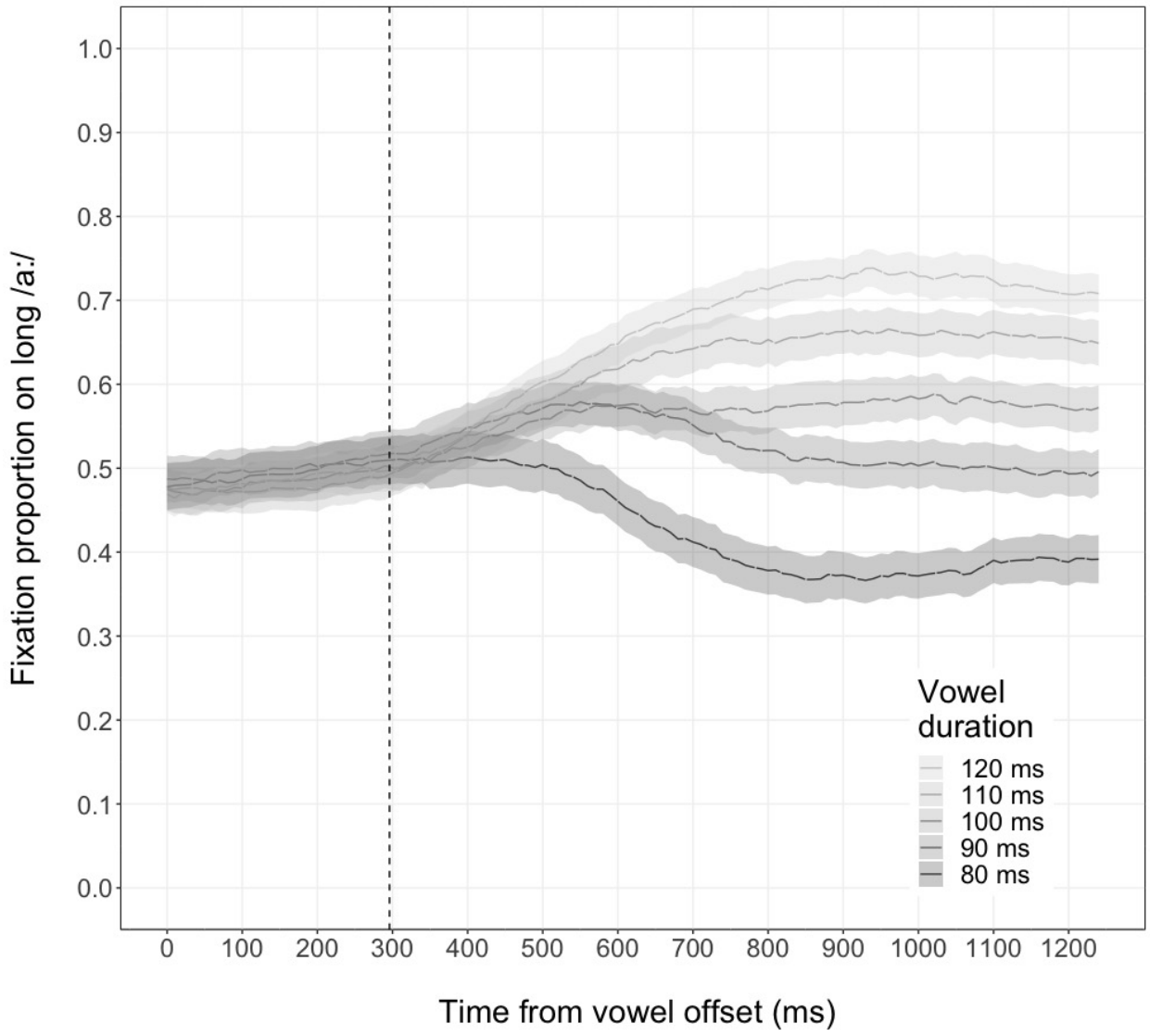

Figure 2: Average fixation proportions to the long /a:/ target word as a function of target vowel duration (80–120 ms), collapsed across contextual speech rate conditions. Increasingly longer vowel durations induced more looks to the long /a:/ target word. Time point 0 is the offset of the target vowel. The gray-shaded areas represent the standard error of the mean. The vertical dashed line indicates the earliest time point (296 ms) at which looking behavior in the 120 ms condition deviated significantly from the looking behavior in the 80 ms condition, as established by the BDOTS analysis.

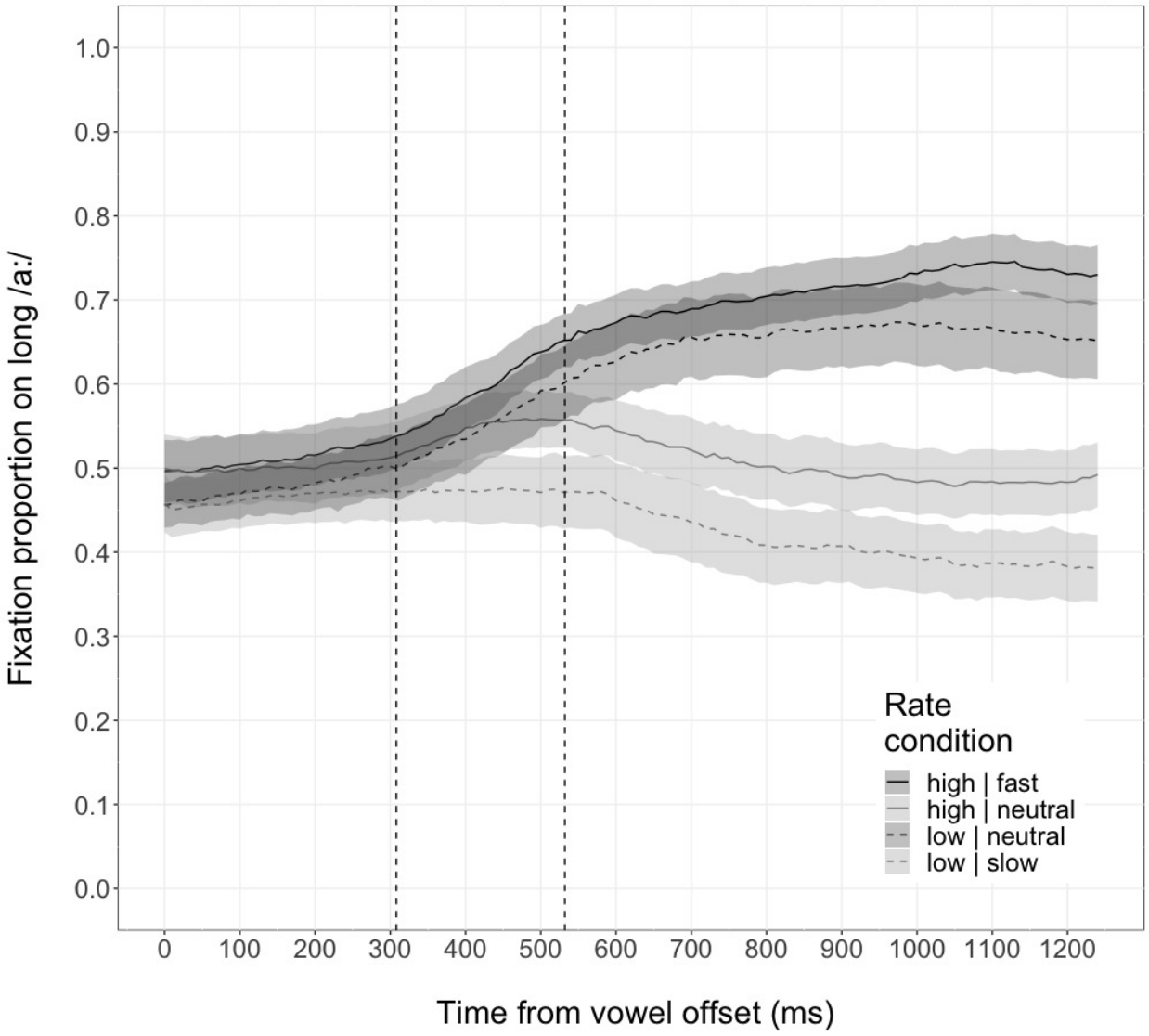

Figure 3: Average fixation proportions to the long /a:/ target word as a function of contextual speech rate, collapsed across vowel durations. Time point 0 is the offset of the target vowel. Line type indicates group. The black lines represent relatively fast speech rates within groups (fast in high-rate group; neutral in low-rate group) and gray lines represents relatively slow rates (neutral in high-rate group; slow in low-rate group). The gray-shaded areas represent the standard error of the mean. 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.

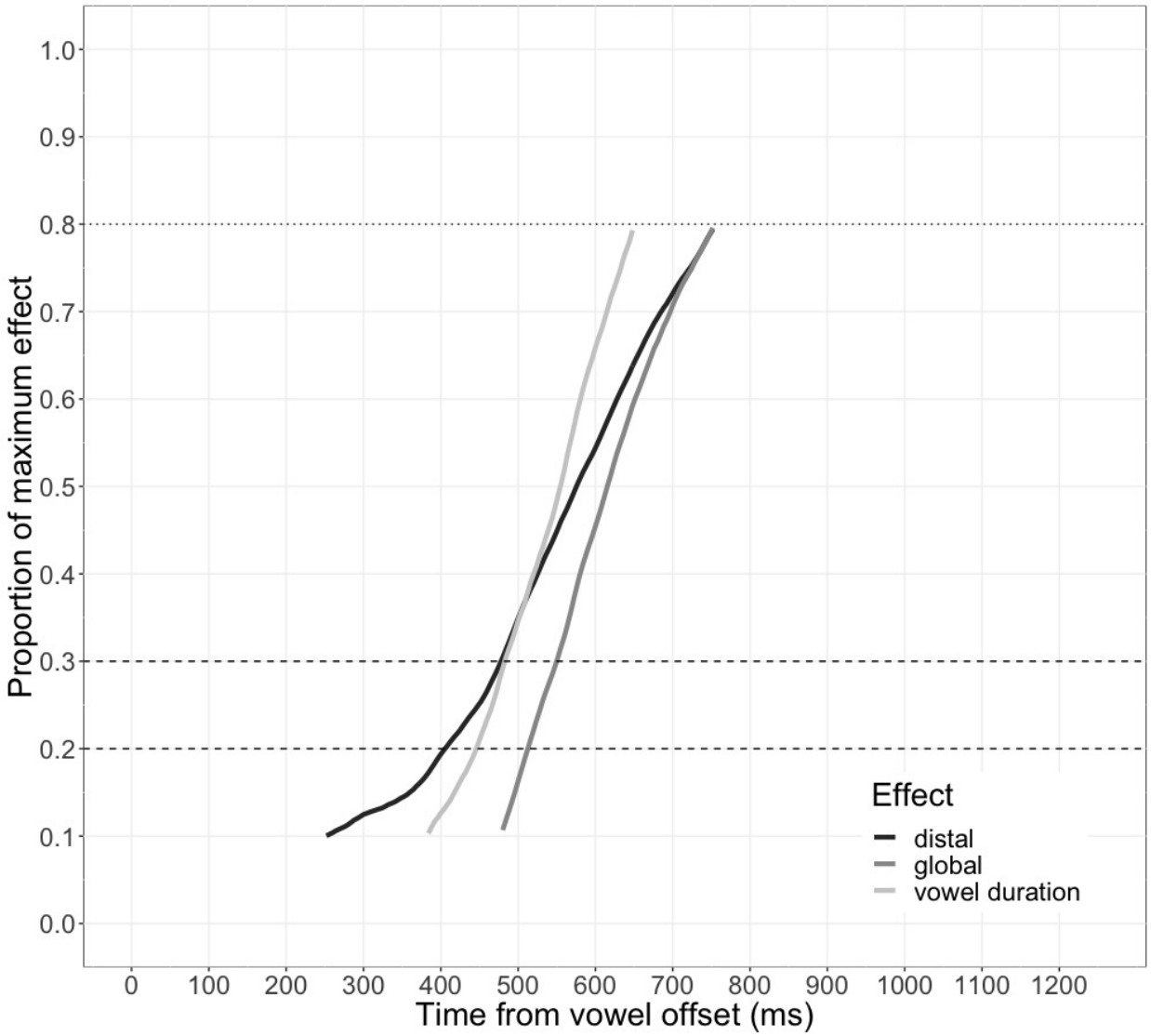

Figure 4: Proportion of the maximum effects for the vowel duration effect (light gray line), the distal rate effect (black line), and the global rate effect (dark gray line). Proportions of the maxima are plotted over time from 200–1000 ms after vowel offset. The dotted horizontal line indicates the time points of 80% of the maximal effect size, at which point the vowel duration effect was significantly different from the distal rate effect in the jackknife analysis (vowel duration = 652 ms; distal = 756 ms; $p = 0.006$). The dashed horizontal lines indicate the time points of 20% and 30% of the maximal effect size, at which points the distal rate effect was significantly different from the global rate effect in the jackknife analysis (20%: distal = 408 ms; global = 514 ms; $p = 0.027$; 30%: distal = 481 ms; global = 550 ms; $p = 0.067$).

List of Tables

Table 1: Results of the BDOTS analyses for when the effects of Vowel Duration, Distal Rate, and Global Rate could be reliably detected.

	Effect Type		
	Vowel Duration (80 vs. 120 ms)	*Distal Rate* (rel. high vs. rel. low)	*Global Rate* (high|neutral vs. low |neutral)
N	32 participants	32 participants	40 items
dropped fits	4/64	3/64	1/80
fitted with AR1	51	54	66
fitted without AR1	9	7	13
autocorrelation t	0.9904	0.9912	0.9994
adjusted α	0.0029	0.004	0.0218
significance region	296 – 996 ms	308 – 996 ms	532 – 996 ms

Table 2: Results of the jackknife analyses for 10%–80% of the maximum value of the vowel duration vs. distal rate effect and the distal vs. global rate effects. F-values and p-values from the ANOVA were adjusted for the repeated use of the data due to the jackknife procedure. Significance is indicated by * for $p < 0.05$, and . for $p < 0.1$.

	Effect Type					
	Vowel Duration vs. Distal Rate			*Distal vs. Global Rate*		
	$F_c(1, 78)$	p	Sig.	$F_c(1, 78)$	p	Sig.
10%	0.09	0.768		0.31	0.581	
20%	0.19	0.667		5.08	0.027	*
30%	0.002	0.968		3.45	0.067	.
40%	0.07	0.794		1.82	0.181	
50%	0.49	0.487		0.89	0.350	
60%	1.72	0.194		0.22	0.644	
70%	2.45	0.122		0.02	0.879	
80%	7.94	0.006	*	0.01	0.943	