Speaking Rate From Proximal and Distal Contexts Is Used During Word Segmentation

Eva Reinisch and Alexandra Jesse
Max Planck Institute for Psycholinguistics

A series of eye-tracking and categorization experiments investigated the use of speaking-rate information in the segmentation of Dutch ambiguous-word sequences. Juncture phonemes with ambiguous durations (e.g., [s] in ‘eens (s)peer,’ ‘once (s)pear,’ [t] in ‘nooit (t)rap,’ “never staircase/quick”) were perceived as longer and hence more often as word-initial when following a fast than a slow context sentence. Listeners used speaking-rate information as soon as it became available. Rate information from a context proximal to the juncture phoneme and from a more distal context was used during on-line word recognition, as reflected in listeners’ eye movements. Stronger effects of distal context, however, were observed in the categorization task, which measures the off-line results of the word-recognition process. In categorization, the amount of rate context had the greatest influence on the use of rate information, but in eye tracking, the rate information’s proximal location was the most important. These findings constrain accounts of how speaking rate modulates the interpretation of durational cues during word recognition by suggesting that rate estimates are used to evaluate upcoming phonetic information continuously during prelexical speech processing.

Keywords: spoken-word recognition, word segmentation, speaking rate, context effects, eye tracking

Speech unfolds over time, so duration is an important cue in spoken-word recognition. Duration can be a segmental cue to phoneme identity (e.g., phonemically long vs. short vowels, voice-onset time duration as a cue to consonantal voicing), or a suprasegmental cue, for example to a word’s lexical stress pattern or to the location of word boundaries (Klatt, 1976). Listeners are sensitive to durational variation and use durational cues such as voice-onset time in a gradual fashion to guide lexical access (e.g., Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002). Durational cues, however, are not perceived in an absolute fashion but rather relative to the rate at which an utterance is spoken. If an utterance is spoken fast, segment durations shorten more it supports a word-initial interpretation (i.e., “eens speer”).

The effect of rate, that is, the effect of a linear manipulation of the acoustic duration of segments adjacent to a target segment, has been investigated extensively in the domain of phoneme perception (e.g., Allen & Miller, 2001; Miller, 1981; 1987; Miller & Dexter, 1988; Miller & Liberman, 1979; Miller & Wayland, 1993). Recently, it has also been demonstrated that rate modulates perception of duration as a suprasegmental lexical-stress cue (Reinisch, Jesse, & McQueen, in press). The present series of experiments investigated speaking rate in relation to another type of durational cue: duration as fine phonetic detail used in word segmentation (e.g., Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). We ask whether linearly manipulated speaking rate influences the interpretation of the duration of juncture phonemes, and hence, where word boundaries are perceived in the segmentation of continuous speech. Furthermore, we ask when during word recognition listeners take rate information into account, and what the relative contribution is of the proximity and amount of rate context.

The focus will be on lexically ambiguous sequences in continuous speech. For example in Dutch, “eens speer” (“once spear”) and “eens peer” (“once pear”) contain the same sequence of sounds but differ in whether the [s] belongs to both words or only the first one. Word boundaries are marked by a variety of acoustic cues that listeners can use (e.g., Cho, McQueen, & Cox, 2007; Mattys, White, & Melhorn, 2005; Nakatani & Dukes, 1977) and duration is an important one of them (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, Liberman, Eccardt, & Pesetsky, 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Burani, & Scott, 1995). The longer a boundary sound is (e.g., the [s] in the example above), the more it supports a word-initial interpretation (i.e., “eens speer”).
Repp et al. (1978) showed that the perceived location of a boundary sound depends on the durations of the following segments. The interpretation of a word sequence as “great ship” vs. “grey chip” depended on the relation between the closure duration of the [t] and the duration of the following [f]. If the duration of the fricative [f] was long, the [t] was assigned to the previous word (i.e., “great ship”) more often than when the frication duration was short. In the present study, we first tested with a categorization task whether the rate of a sentence preceding the boundary sound also influences its perceived location, as Repp et al. predicted. Following a fast rate context, a boundary sound should be perceived as relatively long and thus more often as word initial. Following a slow rate context, the same sound should be perceived as shorter and thus more often as not word initial.

Moreover, because the use of durational cues to word segmentation has been shown to occur during the early phases of word recognition (Shatzman & McQueen, 2006), we also asked whether speaking rate information is also taken into account early in processing or whether it is used later to reinterpret durational cues. Studies on the perception of speaking rate so far mostly used offline tasks such as phoneme categorization or category goodness judgment. These tasks measured the results of the phoneme or word recognition process by asking listeners to give explicit answers about what they perceived. Nevertheless, several studies on the effects of speaking rate suggest that rate effects may indeed occur early during processing. For example, when listeners answered very rapidly in a categorization task, that is, when they initiated their answers before the complete context information could have been processed, then listeners tended to treat following rate context as fast, consistent with the amount of information they had processed at that time (Miller & Dexter, 1988). Further support for the earliness of rate effects in phoneme perception comes from studies that demonstrated the use of rate information from talkers other than the target speaker on the interpretation of the target speaker’s speech (Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000). In addition, rate information of competing speech affects phoneme perception even if the competitor speaker is at a different spatial location from the target speaker (Newman & Sawusch, 2009). This suggests that rate information is taken into account at a prelexical stage of processing, even before early processes such as perceptual grouping of speakers or stream segregation occur.

None of these previous studies, however, directly assessed when during word recognition speaking rate is taken into account. To address this question, our experiments used eye tracking to tap into the online word-recognition process (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). When listening to speech, listeners spontaneously fixate visual referents (including printed words; Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Reinisch, Jesse, & McQueen, 2010) that match their current hypotheses about what is being said. Printed-word eye tracking allowed us to monitor over time how listeners’ hypotheses about upcoming words are modulated and changed over time by preceding speaking rate information.

Using eye tracking, it has previously been shown that listeners use duration as a cue to word boundaries as the speech signal unfolds. When Dutch listeners heard segmentally ambiguous phrases containing the boundary sound [s] they considered an s-initial word “spuit” (“syringe”) longer as a competitor of the non-s-initial target “peer” (“pear”) if the [s] was long than if it was short (Shatzman & McQueen, 2006). The duration of [s] was the only one of several acoustic measures that correlated with eye-movement behavior. We used materials similar to those of Shatzman and McQueen. The critical materials were Dutch segmentally ambiguous word sequences with one of the two boundary sounds [s] or [t]. The question was whether the second word in the sequence (i.e., the target) was [s]- or [t]-initial or not (e.g., “eens peer” vs. “eens speer” and “nooit rap” vs. “nooit trap,” “never quick,” “never staircase”). Target words with the boundary sound [s] always followed the word “eens” (“once,” s-trials, e.g., “wel eens peer,” “wel eens speer”). Listeners were told that “eens” was the intended interpretation of the pre-boundary word and not “een” (“one”). Independent of the interpretation of this pre-boundary word, however, the longer the duration of the [s], the more likely the target word should be perceived as s-initial. Target words with the boundary sound [t] always followed the word “nooit” (“never,” t-trials, e.g., “nooit rap,” “nooit trap”). “Nooit” is not a Dutch word.

Listeners hence had to decide whether the target word started with a [t]. In a sequence of two [t]s at a word boundary, the first one is commonly not released (as was the case for our speaker). In these circumstances, closure duration indicates the number of sounds. The longer the t-closure, the more likely it is that the first [t] has not been released and that the second word is t-initial.

Whereas for [s] the critical information is the duration of the frication noise, for [t] it is closure duration, that is, silence. The presence of a signal in [s] (i.e., frication noise) may be more informative than the absence of an audible signal during the t-closure. When hearing silence, the listener cannot be sure whether indeed silence was presented or whether some acoustic information has been missed. The duration of the frication in [s] may therefore be more easily interpolated relative to the context and lead to larger rate effects than the closure of [t], at least at an early stage of word processing. The use of the two boundary phonemes [s] and [t] thus allowed us to test whether speaking rate effects depend on the type of durational information that should be interpreted relative to rate.

Effects of speaking rate have not been shown with eye tracking before. Reinisch, Jesse, and McQueen (2008) failed to detect such effects and argued that this may have been because their stimuli were not entirely ambiguous. Absolute durational cues (in the case of Reinisch et al. to the lexical stress patterns of words) may have disguised effects of the speaking rate context. We therefore attempted to maximize the chances of detecting an effect of rate here by setting the critical boundary sounds ([s] and [t]) to durations that at a normal speaking rate would be ambiguous between initial and non-initial interpretations. The use of rate information should therefore help listeners recognize the targets. Relative to the rate-manipulated preceding fast or slow context, the ambiguous boundary sounds should be interpreted as long or short. If speaking rate is taken into account during word recognition, then there should be a single mechanism for the interpretation of absolute and relative durational information. Listeners should track rate information in the preceding context and use it to evaluate upcoming durational information optimally during a prelexical phase of processing. This was the hypothesis we tested in Experiments 1 and 2.

Our other major question concerned the relative roles of proximity and amount of rate context. For the interpretation of pho-
neme categories, speaking rate information has been shown to exhibit a stronger effect the closer it is to the target phoneme. This research suggests that speaking rate is calculated within a small time window of about 250 to 300 ms from the target in both directions (Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). Rate context within this time window of approximately one syllable affects the perception of a target sound even if the rate-manipulated material is not adjacent to the target sound (Newman & Sawusch, 1996; Sawusch & Newman, 2000). Here we investigated the relative effects of proximal and distal preceding context. Distal context preceding a target phoneme by more than 300 ms can affect target perception (Summerfield, 1981; Wayland, Miller, & Volaitis, 1994). The effect of this long-range context on phoneme perception appears, however, to be different from the effect of the immediate 250–300 ms context (Wayland et al., 1994). The rate of utterance-length context affects only the location of the phoneme’s best exemplar range but not the width of the best exemplar range. Nevertheless, rate information that is distal to the target can have an effect on target perception.

Previous research suggests that the relative role of distal and proximal rate information is modulated by the rhythmic pattern of an utterance (Kidd, 1989). In an utterance in which all stressed syllables were set to the opposite rate than the unstressed syllables, listeners interpreted the target phoneme relative to the rate of the stressed syllables, even though the stressed syllables were not adjacent to the target (Kidd, 1989). This effect of the non-adjacent stressed syllables was similar in size to effects found in a condition where the complete utterance (i.e., stressed and unstressed syllables) was rate-manipulated. This suggests that the durations of stressed syllables are used to calculate rate. The rate of immediately adjacent unstressed syllables only affected perception when their rate was unexpected with regard to the preceding rate patterns (Kidd, 1989). The relative role of proximal and distal rate context may thus depend on the rhythmic structure of the utterance.

In Experiments 3–5, we examined under which conditions distal rate context affects word segmentation by manipulating both the amount and relative location of the rate context (see Figure 1 and Table 1). To assess the relative importance of distal vs. proximal context, we investigated effects of distal context when the proximal context was informative for the interpretation of the boundary sounds (i.e., when proximal context was rate-manipulated; Experiment 3) and when it was not (i.e., when proximal context rate did not disambiguate the target; Experiments 4 and 5). This allowed us to test whether a distal rate context attenuates the proximal context effect and whether a distal rate context can directly affect word recognition. Moreover, we asked whether a longer distal context (Experiment 4) would yield a stronger rate effect than a short distal context (Experiment 5).

By addressing these questions in both eye-tracking and categorization tasks, we asked whether there are differences in the relative roles of location and amount of rate context in online vs. offline word recognition. Because during word recognition listeners continuously update their interpretation of speech, it seems plausible that they also continuously update their estimate of speaking rate from the incoming information. This need for continuous information update, however, may lead listeners to give most weight to the immediately preceding speaking rate information during word recognition. Therefore, in eye tracking, listeners may be influenced more by proximal than by distal rate context. In contrast, in categorization listeners have more processing time and therefore have the opportunity to post-perceptually re-process speaking rate information from a longer context. Distal rate context in general and the amount of (distal) rate context in particular may thus play a larger role in determining categorization decisions than in influencing eye movements.

**Experiment 1**

Previous research has shown that the duration and hence the location of a juncture phoneme is interpreted relative to the duration of the following phoneme (Repp et al., 1978). Experiment 1 expanded on these findings by examining the influence of an utterance-length preceding speaking rate context on boundary sounds’ perceived location. In a categorization task listeners had to decide whether the target word started with the boundary sound or not. Two different boundary sounds were used: [s] and [t].

The second purpose of Experiment 1 was to find perceptually ambiguous sounds for the subsequent eye-tracking experiments, that is, juncture sounds [s] and [t] which, at a normal speaking rate, are perceived equally often as word-initial sounds or not. Using these ambiguous boundary sounds in Experiment 2–5 should make it easier to observe rate effects, because only the sounds’ interpretation relative to the context can disambiguate their position.

**Method**

**Participants.** Twenty-four participants from the Max Planck Institute’s participant pool took part for a small payment. They were all native Dutch speakers and reported no hearing problems.

**Materials.** Twelve participants categorized two word pairs from the s-trials of the subsequent eye-tracking experiments (“stillen”–“tillen,” “satisfy”–“lift,” and “spear”–“peer,” “spear”–“pear”), and 12 participants categorized two word pairs from the t-trial word set (“toost”–“oost,” “toast”–“east,” and “trap”–“rap,” “staircase”–“quick”). These words were a subset of the words recorded for the eye-tracking experiments (see below). They were recorded in their s- and t-initial forms in the carrier sentences “Ze heeft wel eens x gezegd” (“She once said x”) for s-trials and “Ze heeft nooit x gezegd” (“She never said x”) for t-trials. One token of each carrier preceding the critical sounds was selected for the ambiguous boundary sounds for the subsequent eye-tracking experiments, that is, juncture sounds [s] and [t] which, at a normal speaking rate, are perceived equally often as word-initial sounds or not. Using these ambiguous boundary sounds in Experiment 2–5 should make it easier to observe rate effects, because only the sounds’ interpretation relative to the context can disambiguate their position.

**Figure 1.** Rate and context manipulations across experiments. Examples are given for one word of the t-trial word set. Manipulations for s-trials were identical (see text for details).
The critical sounds were manipulated along 15-step duration continua. To create long continuum endpoints, the boundary sounds in s- and t-initial target words were replaced by representative tokens of [s] and [t] that had a duration one standard deviation above the mean of all recorded s- or t-initial words (147 ms for [s], 138 ms for the [t]-closure). Other steps were created by successively cutting off approximately 8 ms (1/18 and 1/17 of the original durations for [s] and [t] respectively) from the end of the [s] or the t-closure until 33 ms of the [s] and 24 ms of the [t]-closure were left. Other potential segmentation cues were set to fixed values. The closure duration of stops following the [s] in s-trials was set to 80 ms, which was close to its mean duration in the s-initial words. The [t] in the t-trials was chosen such that the RMS-amplitude of the [t]-burst (0.018 Pascal) was within one standard deviation of the average in t-initial and not-t-initial words (0.023 Pascal, standard deviation: 0.007 Pascal).

**Procedure.** For both s- and t-trials each rate context was combined with each word at each step of the continua and repeated six times. This resulted in 540 trials per participant (2 Word pairs × 3 Rates × 15 Steps × 6 Repetitions). The experiment was blocked by word pair with order of pairs counterbalanced across participants. Presentation order within blocks was randomized for every participant with the restriction that all stimuli would be presented once before a repetition occurred. Speaking rates were intermixed at random. Throughout each trial participants saw the two response alternatives on the screen with the s- or t-initial word on the left (e.g., speer peer, trap rap). Auditory stimuli were presented over headphones at a comfortable listening level, 200 ms after the words appeared on the screen. Listeners had to indicate by button press as quickly and as accurately as possible which of the words they heard. The next trial started when a response was given or 3,000 ms from the onset of the critical sounds. Every 45 trials participants were given a short break. The experiment was controlled by NESU software (http://www.mpi.nl/world/tg/experiments/nesu.html).

**Analysis.** Results were analyzed using linear mixed effect models (Baayen, Davidson, & Bates, 2008). A logit link function dealt with the dichotomous dependent variable (i.e., initial response (1) vs. non-initial (0) response). Rate (−.5 = fast, 0 = normal, .5 = slow), step (0–14), and their interaction were entered in the model as fixed factors, and participant and item as random factors. The model maps a factor’s condition with the value 0 on the intercept and assigns regression weights for the adjustments required to map from this condition to every other level of the factor. In all experiments, non-significant interactions were eliminated and the models refitted. Only significant results are reported.

**Results**

Figure 2 shows categorization data in response to the minimal pairs from each trial set.

**S-trials.** For s-trials a main effect of step as well as an interaction between rate and step were found ($b_{step} = 0.33, p < .001$; $b_{step\ rate} = −0.07, p < .001$). More s-initial responses were found at every step of the continuum. No preference for either initial or non-initial responses in a normal rate context was found for steps 5 to 7 ($b_{intercept} = −0.17, p = .46; b_{rate} = −0.11, p = .66$; step 5: $b_{intercept} = −0.05, p = .78; b_{rate} = −0.69, p < .01$; step 7: $b_{intercept} = 0.54, p = .07; b_{rate} = −0.95, p < .001$). The largest rate effect was found for steps 6 and 7 (for 9 out of 12 participants). A duration between these steps (i.e., 86 ms) was selected as ambiguous sound. The vertical line in the left panel of Figure 2 shows the location of this duration on the s-continuum.

**T-trials.** For t-trials overall main effects of rate and step were found ($b_{rate} = −1.29, p < .001$; $b_{step} = 0.55, p < .001$). More t-initial responses were given the longer the t-closure and the faster the context rate. Separate analyses by step suggested that step 5 was perceived as ambiguous at a normal rate context ($b_{intercept} = 0.02, p = .93; b_{rate} = −1.79, p < .001$). An intermediate duration between step 5 and 6 (i.e., 69 ms) was chosen as the ambiguous
sound for t-trials in the eye-tracking experiments. At both steps a large effect of rate context was found (all participants showed an effect at step 5, and 11 out of 12 participants at step 6). The vertical line in the right panel of Figure 2 shows the location of this duration on the t-continuum.

**Discussion**

Listeners use speaking rate information from a preceding sentence in the identification of word boundaries cued by durational information. The faster the rate context, the longer the boundary sound was perceived, and thus the more likely it was interpreted as word-initial. This result held for both juncture phonemes. For s-trials rate effects were evident predominantly at longer steps of the continuum. It is possible that a certain duration of the juncture phoneme is necessary for speaking rate to have an effect. If the absolute duration of the boundary sound is too short, it may be assigned to the pre-boundary word irrespective of rate. Importantly, speaking rate affected the perceived duration of the frication of [s] as well as the duration of silence during the closure of [t].

**Experiment 2**

Experiment 2 introduced a printed-word eye-tracking paradigm to investigate whether and when speaking rate information is used to interpret durational cues during word recognition. As in Experiment 1, the complete context preceding the boundary sounds was rate-manipulated. The demonstration of online rate effects would suggest that listeners take into account not only local durational information but that information from a larger stretch of speech stays available for the interpretation of upcoming information. Such a result would confirm previous suggestions that speaking rate is used at an early stage of speech processing. In addition, the use of eye tracking allowed us to examine the time-course of the speaking rate effect. A categorization task with the same rate manipulation as in eye tracking was included to compare participants’ rate effects in offline and online word recognition. We asked whether the use of rate information during word recognition would be correlated with effects that are measured in a task tapping into the result of this recognition process.

**Method**

**Participants.** Thirty Dutch native speakers from the same population as in Experiment 1 participated. None had taken part in Experiment 1.

**Materials.**

**Eye tracking.** Fifty-four Dutch words of one- or two-syllable length were chosen as targets for s-trials and another 54 words for t-trials. Eighteen target words from each of these sets were s- or t-initial (“initial” condition). The remaining 36 target words in each set were not-s-initial or not-t-initial (“non-initial” condition) but formed a minimal pair with another existing Dutch word starting with [s] or [t]. For example, “peer” and “rap” were non-initial targets, where “speer” and “trap” are also Dutch words (see Appendix). A subset of these minimal pairs was used in Experiment 1 and in the categorization task described below. For the eye-tracking task, displays of four printed words were created for each trial where each target was presented along with a phonological competitor and two distractors. Competitors overlapped segmentally with the target up to and including the first vowel (see Appendix). For initial targets the competitors were non-initial.
(e.g., “peper” and “oester” - “pepper” and “oyster” - for the targets “speler” and “toetsen” - “player” and “test”) and for non-initial targets the competitors started with an [s] or [t] (e.g., “speer” and “tractor” - “pacifier” and “tractor” - for the targets “peer” and “rap”). Words from a minimal pair never occurred together on the screen. Target and competitor sets were matched on their average CELEX lexical frequency (Baayen, Piepenbrock, & Gulikers, 1995; s-initial: \( t(17) = 0.01, p = .99 \); t-initial: \( t(17) = -0.31, p = .76 \); not-s-initial: \( t(35) = -0.75, p = .49 \); not-t-initial: \( t(35) = 1.58, p = .12 \)). Each target-competitor pair was assigned two phonetically, orthographically, and semantically unrelated distractors. Thirty-six filler and six practice displays, each containing four unrelated words, were constructed such that the targets did not begin with [s] or [t] nor formed a minimal pair with an s- or t-initial word.

A female native Dutch speaker recorded multiple tokens of the initial and non-initial targets (x) as well as of their minimal pairs (if available) in the carrier sentences “Ze heeft wel eens x gezegd” for s-trial items or “Ze heeft nooit x gezegd” for t-trial items spoken at a normal rate. Sentence accent was placed on the target word. One token of each carrier preceding the targets and two versions of “gezegd” were selected as the contexts into which all targets were spliced. The initial sound of “gezegd” was voiced in one token and voiceless in the other. This difference is not phonemically distinctive in Dutch but occurs either due to dialectal differences or due to assimilation to the surrounding sounds, as was the case for our speaker. The token of “gezegd” that was used on a given trial was selected so as to maximize acoustic coherence between the target and this following context. In most cases the voiced token of “gezegd” was used after a voiced segment, and a voiceless token after a voiceless segment. The word “gezegd” was always presented at a neutral speaking rate and was thus not informative for the disambiguation of the juncture phonemes. The preceding carriers were chosen such that the selected sentence was of average duration compared to all recorded carriers. Using the same sentence for all items in the respective trial set controlled for variation in pitch contours. The speaking rate of the carrier sentences was manipulated in the same fashion as in Experiment 1 with the exception that here the preceding context was never presented at a normal speaking rate.

The best token of each target word was selected using two criteria. First, to control for speaking rate, the sentence it had been recorded in matched as closely as possible the duration of the carrier sentence that was selected for manipulation. Second, the target token had to fit the selected context in terms of pitch and voice characteristics to minimize audibility of the splicing manipulation. To make the words perceptually ambiguous between initial and non-initial interpretations, targets for the non-initial conditions were created from their s- or t-initial minimal pairs. Vowel-initial words were occasionally produced with creaky voice or delayed voicing relative to the offset of frication. Using these tokens as bases for the stimuli could have produced a bias towards non-initial targets.

Critically, the same ambiguous tokens (one [s] and one [t]) were used throughout the experiment. The ambiguous durations for a normal rate context, as established in Experiment 1, were 86 ms for the [s], and 69 ms for the [t]-closure. The eye-tracking stimuli thus consisted of a concatenation of the rate-manipulated preceding carrier sentences excluding the final [s] or [t] in “eens” and “nooit”, the ambiguous sounds with the durations established in Experiment 1, the target words without the initial [s] or [t], and the following carrier (“gezegd”).

**Categorization.** One word pair of each trial set that was used in Experiment 1 was selected (s-trials: “peer” - “speer,” t-trials: “rap” - “trap”). Their non-initial interpretations were part of the non-initial target condition of the eye-tracking experiment. A subset of steps from the duration continuum presented in Experiment 1 was selected for the categorization experiment here, with a higher density of steps in the ambiguous region. For s-trials the steps used for categorization had durations of 33, 57, 74, 82, 90, 98, 115, and 147 ms (steps 0, 3, 5, 6, 7, 8, 10, and 14). For t-trials these were 24, 41, 57, 65, 73, 81, 105, and 138 ms (steps 0, 2, 4, 5, 6, 7, 10, and 14). The same rate-manipulated sentences as in the eye-tracking experiment were used in the categorization task. Unlike in Experiment 1, the preceding context was never presented at a normal rate.

**Procedure.**

**Eye tracking.** Listeners were positioned 65 cm in front of a 40.5 × 30.5 cm screen. Eye-movements were recorded with an SR Research Eyelink 1000 system at a sampling rate of 1,000 Hz and controlled by Experiment Builder software (SR Research). Each participant heard half of the sentences of each trial type (s-trials and t-trials) in the fast rate condition and half of the sentences in the slow rate condition. Trial types and rate contexts were intermixed and presented in a different random order to each participant. The assignment of words to the four positions on the screen was randomized for each participant with the number of times the target appeared on each position balanced within each condition. Before the start of the experimental list all participants responded to the same six practice trials. Half of the practice items were s-trials, and the other half were t-trials.

On every trial participants saw a fixation cross for 500 ms in the middle of the screen. Immediately afterward, four printed words appeared in the four quadrants of the screen. The center of each word was placed 13.5 cm from the nearest edge of the screen on the horizontal axis and 10 cm on the vertical axis. Words were presented in monospaced, lower case, Lucida Console font, size 20. The auditory stimuli were presented over headphones at a comfortable listening level. The acoustic onset of the sentences was timed such that 1,800 ms after display onset listeners heard the critical sound (i.e., [s] or [t]). Listeners were instructed to click with the computer mouse on the word they heard in the sentence. The next trial started 500 ms after the listeners responded. Every 10th trial a drift correction was carried out. The eye-tracking experiment was immediately followed by the categorization task.

**Categorization.** Each step-rate combination was presented 10 times. The order of trials was randomized for each participant with the restriction that all step-rate combinations were presented before a repetition occurred. Rates were mixed at random. The experiment was blocked by trial type, switching to the word pair of the other type after five repetitions of all steps and rates. Once within each block as well as between blocks participants were given a short break. Order of block was balanced across participants.

On each trial participants saw the words of the minimal pair on the screen, with the s- or t-initial word always shown on the left. After 600 ms, the auditory stimulus was presented over head-
phones at a comfortable listening level. Participants were instructed to indicate which of the words on the screen they heard in the sentence by pressing one of two marked keyboard buttons. The words stayed on the screen throughout the trial until 500 ms after the participant responded or 3,000 ms after the onset of the critical sound. Three hundred milliseconds after either event the following trial was initiated. The categorization tasks were controlled by Experiment Builder software (SR Research).

**Analyses.**

**Eye tracking.** Eye-movement data were analyzed in time bins of 4 ms. Only data that were classified by the eye-tracking system as being fixations were taken into account. A fixation was counted as falling on a word if it was located within a predefined circle of diameter 11 cm around the center of the word. Further, trials containing the not-s-initial target “tillen” and the t-initial target “triest” were excluded from analyses in all experiments. Their competitors differed in vowel length and thus did not show the same amount of phonemic overlap with their targets as the other target-competitor pairs.

The time windows for analyses were chosen from 286 ms to 1,000 ms after the onset of the critical sounds for s-trials and from 269 ms to 1,000 ms for t-trials. These start times were defined by the offset of the critical sounds shifted by 200 ms. A time lag of 200 ms is commonly used as an estimate of the time needed to program and launch a saccade (see, e.g., Hallett, 1986; Matin, Shao, & Boff, 1993), that is, on average, 200 ms after acoustic input eye movements start reflecting the impact of this input on word recognition. The durational information of the critical sounds could have been processed at the earliest only at critical sound offset and thus could potentially only show an effect at that time or later.

Eye movements were analyzed separately for s-trials and t-trials in initial and non-initial conditions using linear mixed effect models (Baayen et al., 2008) as provided in the lme4 package (Bates & Sarkar, 2007) in R (version 2.8.0; The R foundation for statistical computing). The dependent variable was the difference between logistically transformed fixation proportions to targets and competitors (see Barr, 2008). Participant and item were entered as random factors. Speaking rate was entered in the model as numerical fixed factor (fast; .5 = slow). The model maps the hypothetical value of rate = 0 on the intercept and estimates a regression weight with which the values of the factor have to be multiplied to move from the intercept to the respective factor levels. A positive regression weight indicates a greater difference between fixations on target and competitor following a slow than a fast speaking rate context. A negative regression weight indicates a greater difference between fixations on target and competitor following a fast than a slow speaking rate context. P values were based on Markov chain Monte Carlo sampling.

**Categorization.** The categorization experiment was analyzed in a similar fashion as Experiment 1 by using linear mixed-effects models with a logit link function. Because here durations of the continua were not equally spaced, durations of the chosen steps in milliseconds were used as numeric values for this factor, centered on the ambiguous duration that was used for the eye-tracking stimuli (i.e., 86 ms for s-trials and 69 ms for t-trials). The main effect of rate thus referred to the duration of the critical sounds used in eye tracking.

**Results**

**Eye tracking.** Nine trials (0.02%) had to be excluded due to fixations outside the screen. Another 102 trials (2.07%) were excluded because listeners clicked on either the competitor word (51 trials; 1.13%) or outside the defined fixation areas around the words (42 trials; 0.93%).

Figure 3A shows fixation proportions on target, competitor, and the average of the two distractors over time plotted for each rate from the onset of the critical sounds for s- and t-trials in the respective initial and non-initial conditions. The outer two vertical lines mark the time window of analysis. The left inner vertical line marks the end of the last segment that is shared by target and competitor (“target-competitor divergence point”) shifted by 200 ms. The right inner line marks the target offset shifted by 200 ms. Both inner lines represent average values for each condition (see Table 2).

Results of Experiment 2 are summarized in Table 3. For the s-trials an effect of speaking rate was found for s-initial targets (e.g., “speler”). Following a fast rate context, the [s] sounded relatively long and in this way supported the interpretation of the target as being s-initial. In contrast, following a slow rate context, the critical [s] sounded relatively short and thus temporarily supported the not-s-initial competitor (e.g., “peper”) more than the s-initial target (“speler”). For not-s-initial targets (e.g., “peper”), a trend towards a rate effect in the opposite direction from that for the s-initial targets can be seen in the graphs. This effect was in the predicted direction, but failed to reach significance.

For the t-trials, an effect of speaking rate was found in the initial as well as in the non-initial condition. Following a fast rate context the closure duration of the [t] sounded longer, which led listeners to assume a word boundary during the closure. The tendency to fixate t-initial targets (e.g., “toetsen”) more than not-t-initial competitors (e.g., “oeister”) was thus stronger following a fast than a slow rate. As expected, the pattern reversed for not-t-initial targets (e.g., a stronger tendency to look at “rap” than at “tractor” in slow rate contexts).

**Categorization.** Figure 3B shows the categorization data in response to the minimal pairs along the duration continuum following a fast or a slow rate context. Table 4 summarizes the results. For s-trials effects of rate, step, and their interaction were significant. Listeners gave more s-initial responses the longer the duration of the [s] and more s-initial responses were given following a fast than a slow rate context. The effect of speaking rate was stronger the longer the s-durations. For t-trials, only main effects of rate and step were found. Listeners gave more t-initial responses the longer the closure duration and following a fast than a slow rate context. Additional analyses showed that the size of participants’ rate effects in offline categorization did not predict their use of rate information in eye tracking.

**Discussion**

Listeners use speaking rate information to disambiguate word sequences. The categorization results confirm the finding from Experiment 1 that speaking rate affects the perception of durational cues to word boundaries. Critically, the eye-tracking results show for the first time that speaking rate is used for lexical
disambiguation during word recognition. S- or t-initial words were considered as targets more following a fast than a slow rate context, while the opposite was the case for the non-initial targets. Speaking rate affected the perceived duration of the frication of [s] in a similar fashion to the way it affected the perceived duration of silence during the closure of [t].

Overall the time-course of rate effects was very similar to the time-course of effects of boundary sounds that provided only absolute durational cues (Shatzman & McQueen, 2006). As can be seen from Figure 3, speaking rate is used to modulate lexical competition over an extended amount of time. Rate information started to affect word recognition while the acoustic signal of the target was unfolding. The effect, however, extended beyond the end of the words. That is, speaking rate information modulated speech processing even after the targets had been heard in their entirety. This long-lasting target preference based on the rate context is what is likely to be reflected in categorization tasks where listeners respond after target offset (in our task on average 800 ms after the onset of the critical sound). Differences in the time course of competition between initial and non-initial target conditions reflect the delayed support of non-initial targets relative to their initial counterparts. Whereas initial words received support from the onset of the critical sound onwards, non-initial words received support only once the critical sounds had been processed.

Figure 3. Eye tracking (A) and categorization (B) results in Experiment 2. (A) Fixation proportions over time to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines), following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. The outer two vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1,000 ms after critical-sound onset. The inner two vertical lines mark the target-competitor divergence point and the end of the targets shifted by 200 ms averaged over condition (see text for details). (B) Proportion of initial responses for s-trials and t-trials in the categorization task. Duration of [s] and closure duration of [t] were varied on a subset of eight steps of the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment.
In summary, Experiment 2 demonstrated that ambiguous boundary sounds are interpreted in relation to the rate of the preceding context and thus that speaking rate can be used to modulate word recognition. The following experiments further explored the conditions under which speaking rate context is used.

**Experiment 3**

Experiment 3 asked whether the effect of speaking rate in Experiment 2 was based on all preceding rate context or rather only on the immediately adjacent context, as has been suggested in earlier studies (e.g., Summerfield, 1981). To show an influence of distal rate context, we investigated whether a distal rate context that had the opposite rate from the proximal context would attenuate the effect of the proximal context. The proximal context was defined as the words directly adjacent to the target (i.e., “wel eens” for s-trials and “nooit” for t-trials; see Figure 1). The distal contexts were the remaining initial parts of the carrier sentences used in Experiment 2 (i.e., “ze heeft” for both trial types). Although the proximal context differed in the number of syllables between s-trials and t-trials, its absolute duration was similar across trial types (see Table 1).

There are two accounts of how distal rate context could influence effects of the proximal context. An attenuated effect of proximal context relative to Experiment 2 would support accounts that argue for the use of rate information averaged over a time window of approximately 300 ms around the target sound (e.g., Newman & Sawusch, 1996; Sawusch & Newman, 2000). Parts of the distal contexts in our Experiment 3 fall into this time window even for the slow rate proximal contexts (which had a durations of maximally 249 ms) and thus allow for a potential attenuation of the proximal rate effect by distal rate. In contrast, an enhancement of the proximal context effect relative to Experiment 2 would be predicted if the proximal rate was perceived as unexpected relative to the distal rate (Kidd, 1989), that is, the distal context (e.g., slow) should make the opposite rate of the immediately preceding context sound more prominent (e.g., faster) and thus lead to a greater effect on the perception of the critical sounds than if no rate-change had occurred (i.e., as in Experiment 2).

The use of eye tracking and categorization will again allow for a comparison between rate effects in online and offline processing. Distal rate context may have less influence during word recognition than offline. In the online situation listeners may rely on a continuously updated estimate of rate. If so, then that estimate presumably gives more weight to the immediately preceding rate context than to more distal contexts. In categorization, where listeners have time for additional post-perceptual processing, rate from distal context could be reprocessed and thus play a larger role than online.

**Method**

**Participants.** Twenty-six new participants from the same population as in the previous experiments were paid for their services.

**Materials, design, and procedure.** The auditory stimuli of Experiment 2 were manipulated such that the initial portions of the sentences (i.e., “ze heeft” for both trial types). Parts of the distal contexts in our Experiment 3 fall into this time window even for the slow rate proximal contexts (which had a durations of maximally 249 ms) and thus allow for a potential attenuation of the proximal rate effect by distal rate. In contrast, an enhancement of the proximal context effect relative to Experiment 2 would be predicted if the proximal rate was perceived as unexpected relative to the distal rate (Kidd, 1989), that is, the distal context (e.g., slow) should make the opposite rate of the immediately preceding context sound more prominent (e.g., faster) and thus lead to a greater effect on the perception of the critical sounds than if no rate-change had occurred (i.e., as in Experiment 2).

The use of eye tracking and categorization will again allow for a comparison between rate effects in online and offline processing. Distal rate context may have less influence during word recognition than offline. In the online situation listeners may rely on a continuously updated estimate of rate. If so, then that estimate presumably gives more weight to the immediately preceding rate context than to more distal contexts. In categorization, where listeners have time for additional post-perceptual processing, rate from distal context could be reprocessed and thus play a larger role than online.

**Method**

**Participants.** Twenty-six new participants from the same population as in the previous experiments were paid for their services.

**Materials, design, and procedure.** The auditory stimuli of Experiment 2 were manipulated such that the initial portions of the sentences (i.e., “ze heeft”) had the opposite rate than the portion immediately preceding the critical sounds (i.e., “wel eens” and “nooit”, respectively). This was done by exchanging distal con-
texts of the fast and slow carrier sentences from Experiment 2 (see Figure 1 and Table 1). All other aspects of the experiment were identical to Experiment 2. Fast and slow rate were coded for the immediate context.

Results

Eye tracking. Ten trials had to be excluded due to fixations outside the screen (0.26%), and 79 trials (2.03%) for participants clicking outside the target area (53 of these clicks were on the competitor). Figure 4A plots fixation proportions to targets, competitors, and the average of the two distractors over time from the onset of the critical sounds. The graphs suggest that, despite a mismatching distal context, listeners used the rate information in the immediately preceding context to recognize the targets in the initial conditions of both trial sets. Statistical analyses, however, indicate that this numerical trend was not significant (see Table 3). For s-trials and for t-trials, there was no difference between speaking rate conditions in the recognition of either initial or non-initial targets.

Categorization. Figure 4B shows categorization data in response to the minimal pairs along the duration continuum in Experiment 3. As is evident from Table 4, main effects of rate and step were found for s-trials and for t-trials. More initial responses were given following a fast than a slow immediate rate context, and more initial responses were given the longer the critical sounds. Moreover, as indicated by the interaction between step and rate, for s-trials, the effect of rate was stronger the longer the steps were.

Cross-experiment comparisons. The data from Experiment 3 were compared to those from Experiment 2. The additional predictor experiment (levels: Experiment 2, Experiment 3) and its interactions with the other factors were entered into the statistical models. A significant interaction between rate and experiment modulated the information extracted from the distal context. In accordance with previous findings, speaking rate context closer to the target had a greater influence on target perception than distal rate context. In the eye-tracking analyses, modulates word recognition. In addition, main effects of rate were found for both conditions of s-trials. For t-trials, the interaction between rate and experiment was not significant for t-initial and not-t-initial targets. In simpler models only containing main effects, an immediate rate context effect was found for t-trials.

Categorization. Results for the cross-experiment comparisons are summarized in Table 6. For both trial types, main effects of rate and step were found as well as an interaction between these two factors. More initial responses were given following a fast than a slow rate context, and more initial responses were given the longer the critical sounds. The effect of rate was larger at longer durations of the continuum. Critically, for both trial types, the interaction between rate and experiment was significant. The effect of rate was stronger in Experiment 2 than it was in Experiment 3, hence, distal rate context modulated the use of immediate context. For s-trials, the interaction between step and experiment suggests that listeners responded more categorically in Experiment 2 than in Experiment 3.

Discussion

Distal rate contexts influenced the use of speaking rate information that was directly adjacent to the target sounds and thus affected word recognition. The effect of proximal context in Experiment 3 was weaker than in Experiment 2. This is presumably because the proximal context in Experiment 3 was preceded by an opposite-rate distal context, whereas in Experiment 2 it was preceded by a distal context set to the same rate. This attenuation of the effect of immediate context was evident in the categorization task (both trial types) as well as in eye tracking (s-trials). Listeners did not perceive the proximal context as more extreme due to its contrast with the distal context but rather appeared to average the rate information over a longer stretch of time than the proximal context defined here.

Although distal context modulated the effect of the proximal context, listeners nevertheless relied on speaking rate information from the proximal context to interpret the ambiguous sounds. In accordance with previous findings, speaking rate context closer to the target had a greater influence on target perception than distal rate context. In the eye-tracking analyses,

### Table 4
Effects of Speaking Rate, Duration of the Critical Sounds (Step), and Their Interaction in the Categorization Experiments

<table>
<thead>
<tr>
<th>Factors</th>
<th>s-trials</th>
<th>t-trials</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>p</td>
</tr>
<tr>
<td>Experiment 2</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>-1.39</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Step</td>
<td>0.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rate × step</td>
<td>-0.01</td>
<td>&lt;.01</td>
</tr>
<tr>
<td>Rate</td>
<td>-0.43</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experiment 3</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>0.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Step</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate × step</td>
<td>-0.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Rate</td>
<td>-1.19</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experiment 4</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>0.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Step</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate × step</td>
<td>—</td>
<td>—</td>
</tr>
<tr>
<td>Rate</td>
<td>-0.26</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Experiment 5</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate</td>
<td>0.06</td>
<td>&lt;.001</td>
</tr>
<tr>
<td>Step</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Rate × step</td>
<td>—</td>
<td>—</td>
</tr>
</tbody>
</table>

Note. Non-significant interactions that were eliminated from the model are marked with “−”.

---

SPEAKING RATE EFFECTS ON WORD SEGMENTATION 987
the effect of speaking rate did not reach significance, but especially the t-trials showed a numerical trend similar to what was found in categorization.

Experiment 3 demonstrated an influence of distal context indirectly: It showed that distal context that had the opposite rate from proximal context attenuated the effect of the proximal context. Experiments 4 and 5 further explored the influence of distal context. We asked whether listeners would use distal context to disambiguate the critical sounds and whether a long distal context would lead to stronger rate effects than a short distal context.

**Figure 4.** Eye tracking (A) and categorization (B) results in Experiment 3. (A) Fixation proportions over time to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines), following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. Fast and slow rate are coded for the rate of the immediate context. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1,000 ms after the onset of the critical sounds. (B) Proportion of initial responses for s-trials and t-trials in the categorization task. Duration of [s] and closure duration of [t] were varied on a subset of 8 steps from the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment. Fast and slow rate are coded for the rate of the immediate context.

**Experiment 4**

Experiment 4 examined whether distal rate context can directly modulate lexical competition. The context immediately adjacent to the targets was hence set to a normal speaking rate, that is, to the same rate as the target and the following context. Because the critical sounds were ambiguous at a normal speaking rate, the proximal context should be uninformative for word segmentation. In addition, we used longer distal contexts than in Experiment 3 (see Figure 1, Table 1). If listeners use these distal contexts to disambiguate the critical sounds they should be more likely to perceive the targets as initial following a fast than
a slow rate context, and they should be more likely to perceive the targets as non-initial following a slow than a fast rate.

**Method**

**Participants.** Thirty new participants from the same population as before received a small payment.

**Materials.** The same speaker as in the previous experiments recorded a subset of the target words (x) from the s-trials and t-trials in longer context sentences. The new carrier sentences were “Tijdens de lange vergadering heeft ze wel eens x gezegd” (“During the long meeting, she never said x”) for t-trials and “Tijdens de lange vergadering heeft ze nooit x gezegd” (“During the long meeting, she never said x”) for s-trials. Due to Dutch syntactic structure constraints, the sentences used in the previous experiments could not be lengthened by adding more material at their beginnings. One token of each new sentence was selected by matching the durations of the proximal context (“wel eens” and “nooit”) and the target words with the respective durations in the previously used sentences. This ensured comparable speaking rates across experiments. The distal contexts of the selected new sentences were sped up to approximately 66% or slowed down to 133% of their original durations for fast and slow rate conditions. For the normal-rate proximal context, tokens of “wel eens” (199 ms) and “nooit” (181 ms) were taken from the

### Table 5

**Cross-Experiment Comparisons of the Rate Effect in the Eye-Tracking Experiments for S-Trials and T-Trials With Initial and Not-Initial Targets**

<table>
<thead>
<tr>
<th>Factors</th>
<th>s-initial</th>
<th>Not-s-initial</th>
<th>t-initial</th>
<th>Not-t-initial</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 3 vs. Exp. 2 Rate</td>
<td>−4.06</td>
<td>&lt;.001</td>
<td>1.21</td>
<td>.055</td>
</tr>
<tr>
<td></td>
<td>−0.07</td>
<td>.832</td>
<td>−0.27</td>
<td>.380</td>
</tr>
<tr>
<td></td>
<td>1.87</td>
<td>&lt;.001</td>
<td>−0.082</td>
<td>&lt;.05</td>
</tr>
<tr>
<td></td>
<td>0.06</td>
<td>.837</td>
<td>−0.13</td>
<td>.663</td>
</tr>
<tr>
<td></td>
<td>−0.09</td>
<td></td>
<td>−0.09</td>
<td>.737</td>
</tr>
<tr>
<td></td>
<td>−0.11</td>
<td>.684</td>
<td>0.14</td>
<td>.353</td>
</tr>
</tbody>
</table>

Note. Non-significant interactions that were eliminated from the model are marked with “−”.

---

### Table 6

**Cross-Experiment Comparison of Categorization Data**

<table>
<thead>
<tr>
<th>Factors</th>
<th>s-trials</th>
<th>t-trials</th>
</tr>
</thead>
<tbody>
<tr>
<td>Exp. 3 vs. Exp. 2 Rate</td>
<td>−1.38</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−0.11</td>
<td>.381</td>
</tr>
<tr>
<td></td>
<td>0.96</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−0.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−1.19</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.07</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>.457</td>
</tr>
<tr>
<td></td>
<td>0.93</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−0.02</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.42</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.05</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>0.11</td>
<td>.392</td>
</tr>
<tr>
<td></td>
<td>0.01</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−0.62</td>
<td>&lt;.001</td>
</tr>
<tr>
<td></td>
<td>−0.02</td>
<td>&lt;.001</td>
</tr>
</tbody>
</table>

Note. Factors were rate, step, experiment, and their interactions. Non-significant interactions that were eliminated from the model in only s-trials or t-trials are marked with “−”.

---
original recordings. The critical sounds, the target words, and the following context were the same as before.

Procedure.

Eye tracking. The timing of the previous experiments was slightly adapted. Due to the length differences of the new carrier sentences in the fast and slow rate conditions (i.e., 1,169 ms vs. 2,154 ms for s-trials and 1,117 ms vs. 2,065 ms for t-trials; see Table 1), the auditory stimuli were presented at a fixed interval of 300 ms after display onset. Even in the shortest sentence condition (i.e., fast t-trials), listeners still had sufficient time to read the four displayed words before the relevant acoustic information about the targets came available.

Categorization. To keep the categorization experiment with the long context sentences at the same length as the previous categorization experiments, each combination of speaking rate and step of the continuum was repeated 8 times instead of 10 times. The preview time of the response options relative to the start of the auditory stimulus was shortened from 600 ms to 200 ms. The long sentences and the invariant positions of response options on the screen provided participants with sufficient time to prepare their responses.

Results

Eye tracking. Data from 18 trials (0.4%) were excluded due to fixations outside the screen. On 93 trials (2.07%), participants' clicks were outside the target area (67 trials, 1.49%, clicks on competitors; 26 trials, 0.58%, clicks outside the word regions). Figure 5A plots fixation proportions to target, competitor, and the average of the two distractors over time for fast and slow rate from the onset of the critical sounds. Fast and slow refer to the rate of the distal contexts (because the proximal context was set to a normal rate).

Results showed an effect of distal rate context for initial and non-initial targets for the s-trials but not for the t-trials (see Table 3). For s-initial trials, the target-competitor fixation difference was larger following a fast than a slow distal context. As expected, for not-s-initial trials the opposite was the case.

Categorization. Figure 5B shows categorization data for s-trials and t-trials along the duration continuum for fast and slow distal rate contexts. For s-trials, the main effects of rate and step were found (see Table 4). More s-initial responses were given following a fast than a slow distal rate context and more initial responses were given the longer the s-durations. For t-trials, the main effects of rate and step as well as an interaction between these two factors were found. More t-initial responses were given following a fast than a slow distal context and more t-initial responses were given at longer step durations. The effect of rate was stronger for longer closure durations.

Discussion

Experiment 4 showed that distal rate context could directly modulate the perceived duration of the critical sounds. Whereas Experiment 3 showed an indirect effect of distal context by modulating the proximal context effect, in Experiment 4, the proximal context was not informative for the interpretation of the critical sounds. Speaking rate effects thus reflect directly the use of distal context information. This finding was consistent for both types of tasks (although in eye tracking, it was only found for s-trials). Compared to Experiment 3, it thus appears that informativeness of the immediate context is critical to the use of distal rate context in word recognition. The use of distal context was not a result of task strategies. Additional analyses showed that trial number (i.e., trial rank order) in Experiment 4 neither predicted the results nor interacted with the effect of rate. Listeners did not learn during the experiment that informative cues were located further away from the target.

In the eye-tracking task, the distal rate context effects were found for s-trials but not for t-trials. The reason for this may be that prosodic properties of the proximal context “noot” (i.e., in t-trials) were stronger than those of “wel eens” (i.e., in s-trials). In spontaneous speech, “wel eens” is more likely to be reduced than “noot.” Even though sentence accent was only placed on targets, a prosodically strong proximal context (“noot”) may attenuate effects of distal context more than a weak proximal context (“wel eens”) would do. This hypothesis cannot be tested easily because, for example, a shift of sentence accent to the proximal context “wel eens” would still leave “eens” as a weak syllable forming the directly adjacent context. An alternative explanation for this difference between s- and t-trials is that the silence of the [t]-closure may be more difficult to interpret than the frication of [s]. The absence of signal may lead to uncertainty about what had been perceived. The fact that in Experiment 2 rate effects in the eye-tracking tasks were found for both cues, however, does not support this explanation. In addition, the results for both cues were very similar in categorization.

In summary, Experiment 4 shows that listeners use distal context information in the perception of word boundaries. What cannot be decided from Experiment 4, however, is whether the amount of distal context contributed to its effect. Experiment 5 therefore explored whether listeners can also use rate information from a short distal context.

Experiment 5

Experiment 5 investigated the role of the amount of distal rate context on the perception of the ambiguous sound sequences. In Experiment 4, long carrier sentences were used in order to maximize the chances of detecting an effect of distal context. Experiment 5 examined whether a short distal context (i.e., dependent on condition between 218 ms and 617 ms) is sufficient for listeners to disambiguate the critical sounds. In Experiment 5, we combined the short distal contexts from Experiment 3 with the neutral-rate proximal contexts from Experiment 4 (see Figure 1, Table 1). Cross-experiment comparisons then addressed whether a long distal rate context (Experiment 4) would yield stronger rate effects than a short distal context (Experiment 5). Further, a comparison between Experiments 5 and 3 followed up on the question about the role of informativeness of intervening proximal context on effects of distal contexts. Both Experiments 3 and 5 presented the same preceding sentences, and the rate of the distal contexts was manipulated in the same fashion. What distinguished the two experiments was the manipulation of the proximal contexts. Whereas in Experiment 3 the proximal contexts could be used for the disambiguation of the critical sounds, proximal contexts were not informative in Experiment 5.
Method

Participants. There were 30 new paid participants from the same population as before.

Materials, procedure, and design. The experiment was identical to Experiment 4, except that the rate-manipulated distal contexts were taken from Experiment 3. The proximal context, presented at a normal rate, was not informative for the interpretation of the critical sounds.

Results

Eye tracking. Ninety-eight trials (2.18%) were excluded from further analyses due to fixations outside the screen (10 trials, 0.22%) or due to participants’ clicks outside the target area (67 trials, 1.49%, clicks on competitors; 21 trials, 0.47%, clicks outside the word regions). Figure 6A shows fixation proportions to targets, competitors, and the average of the two distractors over time for the two rate contexts from the onset of the critical sounds. Fast and slow rate are coded for the rate of the long distal context. The immediate context had the same rate as the target words. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1,000 ms after the onset of the critical sounds. As evident from Table 3, however, this effect was not significant. A short distal rate context did not modulate lexical competition in any of the target conditions.

Categorization. Figure 6B shows categorization data for s-trials and t-trials along the duration continuum for fast and slow distal rate contexts. For s-trials, the main effects of rate and step...
were found (see Table 4). More s-initial responses were given following a fast rather than a slow distal rate context, and more s-initial responses were given the longer the s-duration. For t-trials, a main effect of step and an interaction between rate and step were found. More t-initial responses were given the longer the t-closure. The effect of rate differed as a function of step. No main effect of rate was present for the duration used in eye tracking.

Cross-experiment comparisons.

Experiment 5 vs. Experiment 4: Amount of distal context. Experiment 5 was compared to Experiment 4 to assess whether a shorter distal context would lead to smaller rate effects than a longer distal context. In both experiments, the proximal context was uninformative about rate. Table 5 summarizes the eye-tracking results. For none of the conditions was an interaction between rate and experiment found. The main effect of rate was significant for s-initial trials. Distal rate context thus affects online word recognition in at least one condition but the amount of distal context appears not to matter.

The categorization data (see Table 6) showed for s-trials main effects of rate and step, as well as an interaction between step and experiment. More s-initial responses were given following a fast rather than a slow rate context, and more s-initial

Figure 6. Eye tracking (A) and categorization (B) results in Experiment 5. (A) Fixation proportions over time to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines), following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. Fast and slow rate are coded for the rate of the short distal context. The immediate context had the same rate as the target words. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the duration that was used in the eye-tracking experiment. The short distal context was rate manipulated. The immediate context had the same rate as the target words.
responses were given the longer the s-duration. Responses were more categorical in Experiment 4 than in Experiment 5. Moreover, the effect of rate was stronger in Experiment 4, that is, the amount of distal rate context influenced the effect of rate on perception. For t-trials, the main effects of rate, step, and an interaction between these two factors were found. More t-initial responses were given following a fast rather than a slow rate context, and the longer the [t]-closure. The effect of rate was stronger the longer the steps of the continuum. Critically, the effect of rate was larger in Experiment 4 than in Experiment 5; hence, in categorization, a longer distal rate context led to stronger influences on listeners’ behavior.

Experiment 5 vs. Experiment 3: Informativeness of proximal context. The comparison of Experiment 5 to Experiment 3 tested whether the informativeness of the proximal rate context affected the use of short distal rate contexts. In Experiment 5, the normal-rate proximal context was uninformative about target interpretation. In Experiment 3, the proximal context had the opposite rate than the distal context and thus could disambiguate the interpretation of the critical sounds. In contrast to the analyses of Experiment 3 where the terms fast and slow rate were used with regard to the proximal context, for the present comparison the data were recoded to reflect slow vs. fast distal contexts.

Results of the eye tracking analysis showed that the informativeness of the proximal context did not modulate effects of distal context (see Table 5). Short distal contexts were not sufficient for listeners to disambiguate the critical sounds. None of the factors or the critical interaction between rate and experiment was significant in any of the conditions.

The comparison of the categorization data (see Table 6) showed for s-trials main effects of rate and step, and an interaction between these two factors. The target was perceived as s-initial more often following a slow rather than a fast distal rate context and the longer the [s]-duration. This context effect is in the predicted direction, as it reflects the use of proximal context in Experiment 3. The slow distal contexts in that experiment were paired with fast proximal contexts, so, in line with all other rate effects with [s] reported here, these fast proximal contexts were followed by more s-initial interpretations. In addition, the effect of rate was stronger the longer the steps of the continuum. Critically, the interaction between rate and experiment was also significant. The effect of rate was stronger in Experiment 3 than in Experiment 5. This result reflects the use of proximal context in Experiment 3.

For t-trials, the main effects of rate and step were found. Participants gave more t-initial responses following a slow rather than a fast distal rate context. As for s-trials, this appears to be because of the use of proximal context in Experiment 3. The critical interaction between rate and experiment was also significant. The effect of rate was stronger in Experiment 3 than in Experiment 5. A three-way interaction between rate, step, and experiment suggests that the effect of rate differed as a function of step in Experiment 5 but not in Experiment 3.

Discussion

Experiment 5 demonstrated that in the categorization task a short distal context can be used to recognize the targets of s-trials and t-trials. Listeners indeed appear to derive rate information from a time window that is somewhat larger than the proximal contexts we used. The assumption of a 300-ms time window over which according to some accounts (Newman & Sawusch, 1996; Sawusch & Newman, 2000) rate information is averaged thus seems to be supported. In comparison to the effect of the long distal context in Experiment 4, however, the short distal context had a much smaller effect. This suggests that either the time window over which rate information is averaged is larger than 300 ms (Newman & Sawusch, 1996; Sawusch & Newman, 2000), or that there is some other mechanism that allows a longer distal context to have a larger effect (e.g., Wayland et al., 1994).

The present series of experiments reveals that effects of the amount of rate context appear to be restricted to offline tasks. Although distal context information can be used during online processing (i.e., eye-tracking results in Experiments 3 and 4) a certain amount of context needs to be available in order to observe such an effect (i.e., no eye-tracking effects in Experiment 5). Additional longer context, however, does not lead to stronger effects than shorter context (see the comparison of Experiments 4 and 5). During word recognition listeners continuously update their speaking rate estimate and thereby use context information that is close to the currently evaluated acoustic cues. In contrast, in offline categorization tasks listeners have additional time to process and even re-process context information before they initiate a response. Therefore little distal context information is sufficient to show an effect (Experiment 5), and a longer distal context shows a larger effect than a short distal context (see the comparison of Experiments 4 and 5).

Concerning the influence of the informativeness of the proximal context on the use of distal rate information, the categorization results confirm the importance of the proximal context in word recognition. Although it seems that the distal context effect was stronger in Experiment 3 (informative proximal context) than in Experiment 5 (not informative proximal context), the direction of the rate effect is crucial. Specifically, in the cross-experiment comparison, the main effect of rate indicated that more initial responses were given following a slow rather than a fast distal context. However, as argued above, this reflects the greater use of (opposite-rate) proximal context in Experiment 3. When interpreted this way, the result is consistent with all other effects observed here (i.e., more initial responses after fast proximal contexts). This suggests that if the proximal context is informative for the interpretation of the critical sounds, listeners use the proximal context rather than the distal context.

General Discussion

A series of categorization and eye-tracking experiments investigated how speaking rate information from a preceding sentence context influences the perception of durational cues to word boundaries. Ambiguous Dutch word sequences of the type “wel eens (s)peer” (s-trials) and “noot (t)rap” (t-trials) were presented in rate-manipulated context sentences such that the boundary sounds [s] and [t] could either be interpreted as pre-juncture phonemes or as the initial sounds of the target words. The juncture phonemes [s] and [t] were chosen to test whether speaking-rate effects depend on the type of durational cue in the signal, that is, frication noise in [s] vs. silence during the closure of [t]. Previous studies on word segmentation showed that the longer the boundary
sounds are, the more likely these sounds are perceived as target initial (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp et al., 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli et al., 2003; Tabossi et al., 1995). Experiment 1 therefore first showed with a categorization task that the duration of boundary sounds and thus their position are perceived relative to the rate of the preceding context. A fast context led to more target initial responses than a slow rate context. Results for the two boundary sounds [s] and [t] were largely similar despite the differences in the acoustic characteristics of their duration cues.

Then two major questions were addressed: First, when during word recognition do listeners use speaking rate information? Second, where relative to the critical sounds must this rate information occur in order for it to be used? Critically, two different tasks provided insight into how speaking rate information modulates lexical competition during word recognition (eye tracking) vs. how speaking rate affects listeners’ responses when more processing time is available (categorization).

Experiment 2 introduced an eye-tracking task to examine rate effects during word recognition. Listeners interpreted durational information in relation to preceding rate context as the signal unfolded. Speaking rate thus appears to be taken into account prelexically to incrementally inform the evaluation of upcoming duration cues. This is in line with previous categorization studies on the use of rate information which already suggested that rate information is used during an early phase of speech processing (Miller & Dexter, 1988; Newman & Sawusch, 2009; Sawusch & Newman, 2000). It also supports other findings showing that word recognition is optimal and incremental (Dahan, Magnuson, Tanenhaus, & Hogan, 2000; Norris & McQueen, 2008; Reinisch et al., 2010, in press; Tanenhaus et al., 1995). The present results thus add direct evidence that, as the speech signal is processed over time, the rate estimate is continuously updated and applied during low-level speech processing.

Previous research on the effects of speaking rate suggested that rate information that is close to the target has a larger influence on target perception than distal speaking rate context (e.g., Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). Experiment 3 supported this literature: When proximal and distal contexts were set to opposite rates, listeners used the rate of the proximal context to categorize the target words. Compared to the uniformly rate-manipulated distal and proximal context in Experiment 2, however, distal context attenuated the effect of proximal context in Experiment 3. Distal context can thus also affect word recognition (see also Kidd, 1989; Summerfield, 1981).

Distal rate context can also directly affect word recognition. In Experiments 4 and 5, the proximal context was set to a neutral rate that was not informative for the interpretation of the target words. Listeners interpreted the targets relative to the earlier distal context (Experiments 4 and 5). The use of distal context in the absence of informative proximal information has also been shown in the domain of spectral context effects on phoneme perception. Spectral characteristics of non-speech precursors, for example, influence listeners’ perception of target phonemes even when thirteen spectrally uninformative tones or 1.3 s of silence intervened between the relevant acoustic context and the target phoneme (Holt, 2005).

Additional results from the present experiments suggest that time-window accounts of rate effects (e.g., that rate is evaluated only within a 300 ms window; Newman & Sawusch, 1996; Sawusch & Newman, 2000) may be too simplistic. In the categorization tasks, a longer distal context led to stronger rate effects than a short distal context (Experiment 4 vs. Experiment 5; see Figure 1, Table 1). The amount of context information thus appears to matter. Similar effects of amount of distal context have been found in a categorization study on the role of distal rhythmic context (i.e., pitch and duration) in the segmentation of compound words (Dilley & McAuley, 2008). Depending on the duration and pitch patterns of preceding context words, listeners grouped the target sequences according to the preceding rhythmic pattern even though acoustically the last three words were kept constant across conditions. Importantly, the effect of context was stronger if it was long (i.e., the full context) than if it was truncated. Listeners thus not only use various types of distal phonetic context information (i.e., spectral information (Holt, 2005); rhythmic information (Dilley & McAuley, 2008), speaking rate here) but the amount of this distal context modulates the strength of the effect (see also Holt, 2006, for effects of amount of spectral context information). In categorization tasks, listeners thus appear to be able to use all previous information, resulting in a cumulative effect of distal context.

Because Experiment 2 demonstrated that listeners immediately interpret upcoming durational information relative to the speaking rate of the preceding context, an important question was how much rate information listeners keep available during online speech perception. Given the need for continuous updates of speaking rate estimates during word recognition, the effect of rate context in eye tracking may not be cumulative (unlike what has been suggested for the categorization task). Indeed, in online processing rate information that is closer to the target appears to be more important than more distal context. The long distal context in Experiment 4 did not show an advantage over the short distal context in Experiment 5. Nevertheless, listeners did use information from the distal rate contexts during word recognition. The effects of proximal context in Experiment 3 were attenuated by the distal context, and, importantly, the long distal context in Experiment 4 affected the interpretation of the boundary sound [s]. It is thus not the case that rate information in distal contexts has no effect on online word recognition. The reason why listeners give less weight to distal context in online than in offline tasks may be the difference in processing time available. During word recognition, as revealed in eye tracking, listeners rely on a continuously updated estimate of rate in resolving the ongoing lexical competition process. In categorization, listeners have more processing time available before a response is initiated. Listeners therefore can re-visit the more distal context for rate information and thus can process the context more thoroughly.

In summary, speaking rate information is used during word recognition to prelexically evaluate durational cues. There were no substantial differences in rate effects due to differences in the acoustic characteristics of these cues. Although proximal context information had stronger effects than distal context, listeners were able to use distal context information to recognize the target words. Effects of distal context were especially prominent when the distal context was long and the proximal context was uninformative for target interpretation. In addition, listeners used distal rate information more when post-perceptual time was available for them to give an explicit response about what they perceived than during
online word recognition. This suggests that, whereas offline tasks such as categorization can reveal the limits of the perceptual system, they do not necessarily show how speech information is used in the early stages of word recognition. With respect to online processing, it appears that local rate information carries the most weight. Listeners efficiently evaluate durational information to word boundaries by immediately taking into account speaking rate information estimated primarily from the proximal preceding context.

References


Reinsch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye-


**Appendix**

**List of Targets and Competitors for S-Trials and T-Trials**

**s-Trials**

**Non-initial targets (target/competitor).** pad/spalk (path/splint), pan/spar (pan/spruce), peer/speen (pear/pacificer), peil/spijt (level/regret), pellen/sperren (pel/blocks), pier/spies (worm/skewer), pil/spits (pill/peak), pit/spin (seed/spider), poel/spoed (pool/hurry), pook/spoor (gearstick/track), port/spons (harbor/sponge), pot/spon (jar/tap), pul/spurt (tankard/spurt), taak/staaf (task/bar), taal/staam (language/stand), taart/taand (cake/standing), tal/stad (number/town), tam/staf (tame/staff), tand/stamp (tooth/kick), tap/stam (tap/trunk), teen/steeg (toe/lane), teiz/stijf (wash Tub/stiff), teken/stelen (sign/steal), tel/stern (count/vote), tempel/stengel (temple/stalk), tikken/stichten (tap/trap), toek/tekenen (stretch/encounter), rem/tred (brake/step), reuzel/reven (lorn/mourn), rib/tril (rib/vibration), ril/trimmen (shiver/exercise), roebel/troetel (rouble/darling), roep/troef (call/trumps), rol/trom (role/drum), rommel/troffel (junk/trump), rooster/troepen (schedule/tropics), rot/trots (horse/pride), rouwen/trauma (grieve/trauma), ruk/truffel (jerk/truffle), urnen/turbo (urns/turbo).

**Initial targets (target/competitor).** slaaf/laars (slave/boot), slank/lap (slender/piece), slepen/lenen (drag/lend), smerig/meten (dirty/measure), speler/peper (player/pepper), spoelen/poes (wash/cat), sleutel/leus (key/slogan), spotten/pols (spot/wrist), schade/chaos (damage/chaos), schelen/chemisch (differ/chemical), sfeer/feest (atmosphere/party), slikken/lichten (swallow/light up), sneeuw/neef (snow/nephew), slim/lig (smart/elevator), smaken/maag (taste/stomach), stress/trechter (stress/funnel), steunen/teugel (support/rein), stekker/technisch (plug/technical).

**t-Trials**

**Non-initial targets (target/competitor).** aak/taai (barge/tough), aal/tanig (eel/tawny), aard/tafel (earth/table), actie/activiteit (action/eighty), al/tak (already/branch), as/tact (ash/tact), eel/teek (horney skin/tick), eer/tevens (honor/besides), egel/tegen (hedgehog/against), elt/term (yard/term), eren/elen (honor/breed), iepen/typisch (elms/typical), il/jil (haste/thyme), innen/tippen (collect/tip), issue/ticket (issue/ticket), oog/toon (eye/tone), oost/toorn (east/rage), rand/trance (edge/trance), rap/tractor (quick/tractor), reed/treema (reason/diariesis), rein/reil (clean/trawl), rek/tref (elasticity/hit), reken/treffen (stretch/encounter), rem/tred (brake/step), reuzel/treuren (lorn/mourn), rib/tril (rib/vibration), ril/trimmen (shiver/ exercise), roebel/troetel (rouble/darling), roep/troef (call/trumps), rol/trom (role/drum), rommel/troffel (junk/trump), rooster/troepen (schedule/tropics), rot/trots (horse/pride), rouwen/trauma (grieve/trauma), ruk/truffel (jerk/truffle), urnen/turbo (urms/turbo).

**Initial targets (target/competitor).** tent/eng (tent/scary), test/engel (test/angel), thee/edel (tea/noble), thema/eeuwig (theme/eternal), tocht/off (journey/victim), toetsen/oester (test/oyster), toeval/oever (coincidence/shore), ton/omgang (cask/contact), tong/order (tongue/order), traag/raar (slow/odd), traan/raad (tear/advice), tragisch/raadels (tragically/riddle), trappen/ramp (kick/disaster), tres/rector (braiding/rector), trend/rechts (trend/right), triest/ring (sad/ring), truc/rum (trick/trump), trui/ruilen (shirt/change).