

# Support for context effects on segmentation and segments depends on the context

Christopher C. Heffner<sup>1,2,3</sup> · Rochelle S. Newman<sup>1,2</sup> · William J. Idsardi<sup>1,3</sup>

Published online: 17 January 2017  
© The Psychonomic Society, Inc. 2017

**Abstract** Listeners must adapt to differences in speech rate across talkers and situations. Speech rate adaptation effects are strong for adjacent syllables (i.e., proximal syllables). For studies that have assessed adaptation effects on speech rate information more than one syllable removed from a point of ambiguity in speech (i.e., distal syllables), the difference in strength between different types of ambiguity is stark. Studies of word segmentation have shown large shifts in perception as a result of distal rate manipulations, while studies of segmental perception have shown only weak, or even nonexistent, effects. However, no study has standardized methods and materials to study context effects for both types of ambiguity simultaneously. Here, a set of sentences was created that differed as minimally as possible except for whether the sentences were ambiguous to the voicing of a consonant or ambiguous to the location of a word boundary. The sentences were then rate-modified to slow down the distal context speech rate to various extents, dependent on three different definitions of *distal context* that were adapted from previous experiments, along with a manipulation of proximal context to assess whether proximal effects were comparable across ambiguity types. The results indicate that the definition of *distal* influenced the extent of distal rate effects strongly for both

segments and segmentation. They also establish the presence of distal rate effects on word-final segments for the first time. These results were replicated, with some caveats regarding the perception of individual segments, in an Internet-based sample recruited from Mechanical Turk.

**Keywords** Speech perception · Temporal processing · Psycholinguistics

People can change the rate at which they speak, and these speech rate changes can affect how speech is perceived. One merely has to imagine a politician announcing an important policy change using the speech rate of an auctioneer to appreciate the extent to which this is true. However, speech rate is not just important in ensuring a political message has appropriate gravitas. Consider a professor in Paris visiting a chalet in Neuchâtel, Switzerland, for a ski holiday. She will suddenly be surrounded by speakers not only of a dialect qualitatively different from the French spoken in Paris but a dialect spoken quantitatively *slower* than the one she is used to (Schwab & Avanzi, 2015). A similar process must occur for an English speaker traveling from Fond du Lac, Wisconsin, to Brevard, North Carolina (Clopper & Smiljanic, 2015; Jacewicz, Fox, O'Neill, & Salmons, 2009), or from Gouda, the Netherlands, to Ypres, Belgium (Verhoeven, De Pauw, & Kloots, 2004). Listeners must adapt to the speech rate of people around them, which can vary from place to place, person to person, situation to situation, or even sentence to sentence.

Speech rate variation can influence the production of the basic building blocks of speech, speech segments. Crystal and House (1988) documented an exhaustive number of changes across different segments for a variety of different speech rates, finding systematic variation in the duration of segments as speaking rate varied. For example, the distribution of voice-

---

✉ Christopher C. Heffner  
heffner@umd.edu

<sup>1</sup> Program in Neuroscience and Cognitive Science, University of Maryland, 1401 Marie Mount Hall, 7814 Regents Dr., College Park, MD 20742, USA

<sup>2</sup> Department of Hearing and Speech Sciences, University of Maryland, College Park, MD 20742, USA

<sup>3</sup> Department of Linguistics, University of Maryland, College Park, MD 20742, USA

onset times (VOTs), the primary cue to word-initial voicing contrasts, differs between fast and slow speech, with slow speech rates corresponding to relatively long VOTs and fast speech rate corresponding to relatively short VOTs (Miller, Green, & Reeves, 1986). Even second-language speakers generally hold to these relations between speech rate and the timing of individual phonetic landmarks (Bent, Bradlow, & Smith, 2008; Schmidt & Flége, 1995).

These changes in rate require that listeners adjust their perception accordingly. If they do not, they may run the risk of misperceiving voiceless stops produced at a fast rate as if they were voiced (and voiced stops in a slow context as voiceless). A number of studies have shown that the duration of adjacent (i.e., *proximal*) phonemes and syllables can alter listeners' perception of potentially ambiguous sounds, and not just for contrasts that rely on VOT. The perception of word-final stop voicing, for example, is primarily derived from the duration of the vowel immediately preceding those stops (Raphael, 1972). The perception of word-initial stops, meanwhile, can be influenced by the duration of following vowels, even when listeners' responses are speeded to a point at which they likely could not have extracted all of the vowel's duration information before responding (Miller & Dexter, 1988). Proximal duration information additionally modulates the perception of internal category structure (Miller & Volaitis, 1989; Volaitis & Miller, 1992).

Most studies of speech rate cues in perception have focused on the speech rate of proximal syllables. Although some early studies suggested that speech information far away in time from an ambiguous region of a sentence (i.e., the *distal* context) can also change what is heard within the ambiguous region (Miller & Liberman, 1979), evidence for these distal effects on segments is equivocal at best. In one study, distal speech rate influenced the perception of which segmental token was the “best” example of a category more than it did which tokens were perceived as belonging to which category (Wayland, Miller, & Volaitis, 1994). In another study, distal speech rate only had an effect when modified in certain rhythmic patterns across the context (Kidd, 1989). However, many of those studies manipulated proximal and distal context simultaneously. This made it impossible to disambiguate the effects of the rate of individual adjacent syllables from the speech rate further away within a sentence (Miller & Grosjean, 1981). Some studies even suggested that many of the rate effects found were the result of artifacts caused by stimulus creation (Shinn, Blumstein, & Jongman, 1985). Even experiments that used similar methods but were designed to disambiguate proximal and distal speech rate effects largely and sometimes exclusively saw proximal effects taking priority (Summerfield, 1981).

As such, many later studies of the effects of speech rate on segment contrasts were actually more complex extensions of the effects of proximal rate alone, thus putting aside the

possible contribution of distal rate effects. Listeners were said to compute ratios between the duration of individual segments and the duration of adjacent ones (Boucher, 2002). Perhaps some of the most thorough studies to investigate the effects of distal speech rate on segments have been performed by Newman and Sawusch. Newman and Sawusch (1996) examined a wide variety of segmental contrasts—/t/ (“ch”) and /l/ (“sh”); /t/ and /s/; /b/ and /p/; and /d/ and /t/—found, repeatedly, no effects of distal speech rate context on the perception of segments. They hypothesized that listeners could only incorporate acoustic information about speech rate into their judgments of segments within a certain temporal window, perhaps 300–400 milliseconds in length. These results were maintained when speech rate information was found after the point of acoustic ambiguity (rather than before it) or was uttered by another speaker. Although proximal effects were observed, there was no support for the idea of distal rate effects on segments (Sawusch & Newman, 2000). Overall, evidence for distal effects on segments has been inconsistent, and, when present, weak. The speech rate effects that have been found may be constrained by a discrete period of time in which rate cues may be used.

The spotty evidence for distal speech rate effects on segments differs quite markedly from recent research examining the effects of distal contextual information on word segmentation. Here, there is evidence for strong distal effects. One group of studies along these lines has involved the use of lexically ambiguous syllable sequences such as *down-town-ship-wreck* or *gang-ster-ling-go*. Given enough acoustic ambiguity, these can be parsed as sequences ending in a disyllabic word (*downtown shipwreck* and *gangster lingo*, respectively) or as sequences ending in a monosyllabic word (*down township wreck* and *gang sterling go*, respectively). What is interesting is what happens when these sequences are preceded by additional, unambiguous lexical information. The prosodic expectations established in the context—factors such as duration and pitch patterns—then carryover to the ambiguous syllables, influencing their segmentation. For example, if the first syllable in the strings above (*down* and *gang*, respectively) is lengthened to roughly match the duration of individual words in the preceding context, the following syllables are parsed in line with the monosyllabic word standing on its own (as, say, *down township wreck*; Dilley & McAuley, 2008). These distal prosodic patterns take precedence over semantic cues to word boundaries (Dilley, Mattys, & Vinke, 2010) and can be indexed by various word-segmentation-associated ERP components (Breen, Dilley, McAuley, & Sanders, 2014).

These effects can also be observed in more natural sentential contexts. Dilley and Pitt (2010) were the first to establish the effects of distal speech rate on the perception of function words (grammatical words such as *are*, *or*, and *her* that provide grammatical information to the listener). Crucially, these

words are often acoustically reduced, perhaps in proportion to their predictability or frequency, which means that there is often acoustic ambiguity to their presence in fluent, casual speech (Bell, Brenier, Gregory, Girand, & Jurafsky, 2009). For example, the function word *or* in the sentence *Anyone must be a minor or child* can be pronounced [ə] (“er”), which is identical to the [ə] sound in the final syllable of *minor*. In fluent speech, these two sounds are coarticulated and can merge together, making it unclear whether there is a single longer [ə] or two shorter [ə] sounds in a row. Thus, there is acoustic ambiguity to the existence of the word *or*; the [ə] of *or* might just be a continuation of the [ə] in the previous syllable.

What Dilley and Pitt (2010) found is that the speech rate of the distal portion of the sentence can exert an influence over the segmentation of the part of the sentence with acoustic ambiguity. By slowing down the speech rate of portions of the phrase such as *Anyone must be a mi-*, without changing anything in the immediately adjacent context to the ambiguous vowel, listeners went from transcribing a word like *or* (i.e., hearing a word boundary within the ambiguous region) approximately 80% of the time to transcribing it approximately 30% of the time. This is much stronger than any of the distal speech rate effects observed in the segment literature, to the extent that there are effects in the first place. In eye-tracking studies, these results emerge quickly after the onset of the point of ambiguity (Brown, Salverda, Dilley, & Tanenhaus, 2011, 2015). And there is emerging evidence that distal prosodic cues affect the perception of word segmentation in languages besides English, such as Russian (Dilley, Morrill, & Banzina, 2013), Dutch (Reinisch, Jesse, & McQueen, 2011), and Mandarin (Lai & Dilley, 2016).

It may be tempting to conclude from this review that there is something fundamentally different about the perception of segments and the perception of segmentation that explains the differences between the two types of acoustic ambiguity in the strength of distal speech rate effects. However, the type of ambiguity being studied was not the only difference between these studies. Across experiments, the researchers who have studied each phenomenon have tended to use different methodologies. It is possible that these differences in methods drive the contrast between the findings in the segmental perception and word segmentation literatures.

Experiment 1 of Newman and Sawusch (2009) and Experiment 1 of Dilley and Pitt (2010) provide a useful contrast to illustrate this point. Newman and Sawusch were interested in the influence of the rate of a carrier sentence on the perception of an ambiguously voiced initial stop token in the nonword [kajp~gajp] (“kipe” or “gipe”). This was created from multiple recordings of a single carrier sentence by two speakers, with speech rate being manipulated naturalistically (i.e., the speakers were told to speak quickly or slowly). Participants in the study were told to listen to the sentence

and rate the last word on a Likert scale for how good an example of each possible initial consonant it was. Dilley and Pitt (2010), meanwhile, examined the influence of distal speech rate on an ambiguous word segmentation task, as previously mentioned. They used the recordings of 50 different experimental items by 12 different speakers; these recordings were then manipulated artificially to create normal and slowed versions of each item. Participants in the study were told to listen to the sentence and write down the entire sentence after it finished playing without any attention being drawn to a particular region of the sentence. This suggests several possible methodological differences that could have contributed to the differences in the effects observed: for example, the number of speakers in the experiment, the speech rate manipulation, the lexicality of the target items, the participant response, or, in other studies in the literature, the time course of what was considered “distal.”

In the studies reported here, the methodological disparities present in previous studies were much better matched. Segment and segmentation contrasts were set up to be similar to each other. Standardizing the methods allowed for examination of whether the differences reported in each study were the result of methodological concerns alone or whether there is something fundamentally different between segmentation and segments. To do this, clusters of stimuli were created that were identical except in certain critically ambiguous regions, and, even then, were fairly similar across conditions. These clusters included two pairs of sentences: one with ambiguity to a segmental voicing contrast, as in *The merchant sold Canadian coats* and *The merchant sold Canadian goats*, and one with ambiguity to word segmentation, as in *The merchant sold Canadian oats* and *The merchant sold Canadian notes*. The segment contrasts that were used in this experiment involved consonant voicing, both word-initial and word-final. The word-initial voicing ambiguities were of a standard type in the segment literature. A relatively long voice onset time (VOT) in this case would lead to the perception of a voiceless token (as in *Canadian coats*). A relatively short VOT would lead to the perception of a voiced token (as in *Canadian goats*). Word-final segment pairs, such as *Bailey has much beet/bead knowledge* (matched with *Bailey has much bee/bean knowledge*, which was segmentation-ambiguous), were also employed, which differed primarily in the perceived length of the vowel immediately preceding the word-final stop.

Each pair of segment-ambiguous items was complemented by a segmentation-ambiguous pair to form a cluster. For the *Canadian coats/goats* example, the matched pair was *The merchant sold Canadian oats* and *The merchant sold Canadian notes*. These sentences were ambiguous in the segmentation of the /n/ sound between *Canadian* and *notes*. If the /n/ sound was perceived as quite short, a listener might posit a word boundary at the end of the sound, thus segmenting the

phrase as *Canadian oats*. However, if a listener perceived the /n/ sound as long enough to sustain two /n/ phonemes, the listener might then segment the phrase as *Canadian notes*. Like in some of the studies of Dilley and colleagues, this particular ambiguity type involves an ambiguity in the number of adjacent identical segments (in the number of /ə/ sounds for *minor or child*, the number of /n/ sounds for *Canadian notes*). However, unlike in those previous studies, this does not lead to a reduction in the number of words perceived, but instead in the lexical content of the utterances in question. This type of ambiguity therefore bears more resemblance to those found in previous studies of stops that straddled syllable boundaries (Fujimura, Macchi, & Streeter, 1978; Repp, 1978; Schouten & Pols, 1983). In contrast to these studies, however, which often used artificial contexts (for example, embedding the contrasts within two carrier vowels) where it was uncertain which type of prosodic boundary that participants perceived, the present study involved fully specified lexical contrasts. Furthermore, rather than stops, ambiguous nonstop contexts (/s/, /n/, /l/, etc.) were employed with the reasoning that truly ambiguous stops would be harder to elicit from speakers naïve to the purposes of the study than these consonants. Thus, the closest analogy to this particular segmentation ambiguity is Reinisch et al. (2011), who examined the effects of distal speech rate on Dutch segmentation contrasts such as “*eens (s)peer*” (once (s)pear), with an ambiguous /s/ potentially straddling a word boundary. If the past literature is borne out here, the effects of distal context should be stronger for this segmentation contrast than for the segment contrast.

Besides the creation of stimulus clusters that are roughly equivalent across segmentation and segments, a number of other aspects of the experimental context were held constant across the study. Sixty different experimental items produced by six different speakers were recorded. Participants had to write down the last two words that they heard after every sentence. And the manipulation was artificial, not naturalistic, so speech rate alone could be probed, rather than other, co-varying characteristics of fast or slow speech (Adank & Janse, 2009; Crystal & House, 1988). Three different speech rate conditions were used: one with speech rate as originally recorded, and two with progressively slower distal context rates.

Four explanations for the differences between segmentation and segments are considered here. The first considered here, and perhaps the closest to our initial hypotheses, relates to the way that each phenomenon is represented. Word segmentation is generally described as a suprasegmental contrast. The location of word boundaries is something that is generally said to be predicated on levels of representation that include syllables and other higher aspects of sound-based structure, depending on the particular prosodic hierarchy one assumes (Shattuck-Hufnagel & Turk, 1996). This requires taking in information from a larger time course than just a single segment alone. Segment contrasts do not require these larger

constructs to be perceived, with perhaps only adjacent contexts being informative for most segmental distinctions. Perhaps word segmentation can be more strongly affected by context because word segmentation ipso facto requires information distributed across the context. For instance, Poeppel and colleagues have proposed two separate, concurrent streams of phonetic processing, one with a short time window (perhaps 20–50 ms), appropriate for segmental processing, and another with a long time window (perhaps 150–300 ms), appropriate for syllabic processing (Boemio, Fromm, Braun, & Poeppel, 2005; Hickok & Poeppel, 2007; Poeppel, Idsardi, & van Wassenhove, 2008). Under this account or from some other origin, it is possible that distal speech information is simply “less distal” in a mode of processing that characterizes suprasegmental processing than in a mode of processing that characterizes segments. The representations of segment and segmentation information may therefore drive the differences in the use of distal speech rate cues. Segmentation ambiguities should under this explanation be affected by distal rate for both word-final and word-initial contrasts, regardless of the exact context modified, while segment ambiguities should not be affected by distal rate.

A second possible explanation involves the time course of segment and segmentation processing. It may simply take longer to commit to any particular segmentation of an utterance than it takes to pick up on a segmental contrast. This longer processing time allows for a listener to pick up more information, including distal speech rate cues, when perceiving any particular word segmentation ambiguity. This type of explanation may receive some support from what have been termed lexical theories of word segmentation, wherein word boundaries are only posited after segmental perception is complete, perhaps in part through a time lag built into the system (Mattys, 1997; Mattys, White, & Melhorn, 2005). Examples of these types of theory include TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994), and Shortlist B (Norris & McQueen, 2008). Under these approaches, word segmentation requires accurate segment information to proceed. As such, the differences in distal speech rate effects found here would essentially be an accident of the time course of processing; word segmentation would allow more temporally distributed information to affect it because it just happens to have more widely distributed information available to it. That is, all information is used that is available at the time of a decision; decisions just happen to be made later for segmentation. Such an explanation would also predict rate effects on segmentation to be larger than rate effects on segments. This explanation would be challenging to disambiguate from the first without real time information (as comes from, say, eye tracking), as its predictions in terms of after-the-fact responses are very similar to the previous one.

A third potential explanation for the differences found between the previous studies relates to the position of the

segment ambiguities within a word. Essentially without exception, studies of distal rate effects on segments have involved word-initial ambiguities, and typically word-initial voicing ambiguities. Although quantifying the location of these effects for segmentation studies is more challenging (as changes in segmentation can sometimes lead to changes in the position of a certain segment within a stream of words), the literature there is more mixed. Although the debate remains unsettled, some recent work has argued that the idea of rate adaptation on word-initial voicing is actually unnecessary after taking into account word frequency and other factors (Nakai & Scobbie, 2016). There is no reason, though, that word-final ambiguities would be immune from the influence from the rate of the distal context. Vowel duration, like any other duration, could be perceived relative to the length of the phonetic information found in the context of the vowel. This position-based explanation would suggest that, even if word-initial segment ambiguities are not subject to distal influence, word-final ambiguities might. Effects might emerge for word-final tokens that would not be present for word-initial ones. Such a possibility would be in line with neither representational nor processing-based accounts, as both such accounts would predict that word-final voicing would be no more influenced by distal rate than word-initial voicing.

A final potential explanation for the differences between studies of distal rate effects on segments and on segmentation is in the definition of *distal*, or context, when considered with regard to speech rate. Dilley and colleagues have generally relied on a definition of distal speech rate as being more than one syllable removed from a point of possible ambiguity. However, depending on the syllables used, this may be still well within the 400-ms temporal window that Newman and Sawusch (1996) considered *proximal* in their studies. It is entirely possible that the erstwhile distal rate effects found in the studies of word segmentation would just be the result of information very close to the ambiguous word boundary, but not actually adjacent, as has been argued in the segmental literature (Newman & Sawusch, 1996; Summerfield, 1981), particularly because some studies have shown changes in the size and scope of effects depending on what, exactly, is considered *distal* (Kidd, 1989; Reinisch et al., 2011). For example, Reinisch et al. (2011) used four different distal rate configurations to evaluate rate effects in Dutch. In two of their conditions, distal rate was manipulated without manipulation of the proximal rate, with the only difference being in the amount of lexical material in the distal context. Participants were generally no more likely to hear [per], “pear”, rather than [sper], “spear” in rate-modified conditions with a long distal context than a short one. In another condition, the proximal and distal rates were manipulated in opposition to one another; in this condition, participants’ reliance on proximal rate information was attenuated by the speech rate information in the distal context. Finding a relationship between distal and

proximal rate information spotlights the importance of testing both distal and proximal effects, and varying context durations, in studying what rate information listeners use when understanding the speech signal.

In this study, four different contexts are considered to evaluate the strength of distal rate effects across different context definitions. For Experiment 1, Experiment 2, and Experiment 3, three different definitions of distal context were employed across participants, varying in terms of how “close” or “far” in time the distal context was to the critical contrast. These definitions were chosen to allow directly comparison of the methodological choices made in different prior experiments. One definition, used in Experiment 1, matched that of Newman and Sawusch (1996) in that it included everything more than 400 ms away from the target. Another, used in Experiment 2, accorded better with the specific temporal window used by Dilley and Pitt (2010) in that it included everything more than one syllable away. A third one, used in Experiment 3, involved manipulating the difference between the two previous definitions; that is, the region from one syllable out to 400 ms out. These conditions are described in greater detail in each experiment’s respective Method section. If the definition of “distal” affects the strength of distal rate effects, it should be expected that the strength of distal rate effects should vary depending on the experiment. Based on Reinisch et al. (2011), it would be expected that the more liberal definition of distal used in Experiment 2 should lead to stronger distal effects than the more conservative one of Experiment 1. Experiment 4, which involves the examination of proximal rate effects, was intended as a control experiment.

A final issue of interest was to compare data collected from a typical university population to data collected from Amazon’s Mechanical Turk crowdsourcing service. Mechanical Turk allows corporations and other interested parties to divide up non-automatable work across many human workers. For researchers in the behavioral and social sciences, it also provides convenient and easy access to a broad pool of participants—in many cases, a more diverse and representative sample than is easily available on a college campus (Buhrmester, Kwang, & Gosling, 2011). Many results obtained through the use of Mechanical Turk are in line with results obtained from samples in person, despite some of the demographic differences that exist in the populations (Goodman, Cryder, & Cheema, 2013). Within the language sciences, Mechanical Turk has most frequently been used for purposes like speech transcription (Marge, Banerjee, & Rudnicky, 2010) and norming and acceptability data for syntactic experiments (Sprouse, 2011).

Only a handful of studies have used Mechanical Turk participants for tasks in speech perception. This is understandable given worries about the audio equipment available to Mechanical Turk participants. The studies that do exist have generally compared Mechanical Turk participants to

participants collected before the advent of Mechanical Turk (Kleinschmidt & Jaeger, 2012) or have not used a comparison group at all (Kurumada, Brown, & Tanenhaus, 2012). However, the idea that Mechanical Turk participants may be useful for experiments in phonetics has recently gained some support from a careful validation of Mechanical Turk of previous language-related experiments using word identification and lexical decision tasks (Slote & Strand, 2016). Since there was an opportunity to use participants from Internet-based and live groups in this study, they were used to explore whether participants recruited through Mechanical Turk would show the same effects that were found in the lab. The belief was that they should, despite an expected increase in the variability in participant backgrounds and demographic characteristics.

## Common methods

### Materials

Sixty stimulus “clusters” were constructed that were designed to keep as many attributes of the context equivalent as possible across conditions comparing distal effects on segments to distal effects on word segmentation. They are listed in the Appendix. These clusters were composed of two pairs of stimuli that were ambiguous either in terms of segmentation or in terms of segment identity. Pairs with word segmentation ambiguity differed in the segmentation of a continuant segment (in the study, [n], [m], [s], [z], [f], [v], [θ], [l], and [ɹ]) between two words. For example, the sequences “Canadian oats” and “Canadian notes” differ only in whether a word boundary is placed after the phrase-medial /n/ (as in “Canadian oats”) or in the middle of the phrase-medial /n/ (as in “Canadian notes”). In “Canadian oats/notes,” the point of primary interest is in whether the second word started with an /n/, but this was not always the case; the primary point of ambiguity in the pair “bee knowledge” and “bean knowledge” was whether the first word ended with an /n/. Pairs with segment ambiguity differed in the voicing of a phrase-medial stop consonant. An example of a pair in this condition is “Canadian coats” and “Canadian goats,” where the second word of the former phrase starts with a voiceless consonant while the second word of the latter phrase begins with a voiced one. Again, as with the word segmentation ambiguity, pairs could also differ in word-final voicing: “beet knowledge” and “bead knowledge” provide an example of such a pair.

A pair of items with word segmentation ambiguity was matched to a pair with segment ambiguity to create a cluster. For example, “Canadian oats/notes” was paired with “Canadian coats/goats,” while “bee/bean/beet/bead knowledge” similarly formed a cluster. Wherever possible, clusters were created where the only difference across ambiguity type was in the critical consonant that indicated

participants’ word segmentation or voicing perception, as with the clusters given above. However, this was not always possible, as there are few English tetrads that have appropriate phonetic properties. In these circumstances, pairs were clustered that shared other properties and could be embedded in similar lexical contexts; for example, “Pat’s car/scar/card/guard” formed a cluster, as did “bar/barn/dock/dog nearby.” These clusters were then embedded in identical lexical contexts, to form clusters such as “The merchant sold Canadian oats/notes/coats/goats.” and “The tornado threatened the bar/barn/dock/dog nearby.” In each case, the two words making up the critical phrase (underlined in the previous examples) were placed at the end of the sentence, with four to seven syllables of prior context. The sentences were grammatical, if occasionally implausible, under each possible interpretation. Fillers were also created that matched the approximate number of syllables within the sentence and in their semantic plausibility (or lack thereof) but did not have the key ambiguities of the experimental items.

Six native speakers of American English (three male, three female), who were naïve to the purpose of the study, recorded each set of clusters as well as the fillers. There was some time pressure placed on speakers to encourage more casual pronunciation. Trials where this led speakers to be cut off while producing sentences were discarded, as well as tokens with speech errors and with discrete pauses between repeated critical segments in word segmentation ambiguities (e.g., between each /n/ in “Canadian notes”). A variety of acoustic measurements were taken in an attempt to determine the clusters most similar to each other. In the end, tokens were chosen with approximately equal sentence context durations. These tokens were selected and used for further analysis.

For each item within each cluster, additional measurements were taken to create the experimental stimuli for this study. In particular, for sentences with ambiguity to word segmentation, the duration of the critical segment (e.g., /n/ in “Canadian oats/notes”) was measured, which is the primary cue to the location of the word boundary (see, e.g., Shatzman & McQueen, 2006). A critical duration for the sentences with segment ambiguity was also measured. This differed between word-initial and word-final positions. For word-initial pairs like “Canadian coats/goats,” the voice onset times for the critical word-initial stops served as the critical duration, as that is the primary cue to the voicing of the segment (Lisker & Abramson, 1967). For word-final pairs such as “beet/bead knowledge,” the duration of the immediately previous segment (e.g., /i/ in “beat” and “bead”) was the critical duration, as it is the primary driver of word-final voicing stop contrasts (Raphael, 1972).

The duration of critical segments across the different members of each pair were compared to create maximally ambiguous versions of each recording: the amount of ambiguity to the location of the word boundary for segmentation-

ambiguous sentences or the voicing of the critical consonants for segmentally ambiguous sentences. On average, the duration of the critical consonant in single-consonant version of the sentences (“Canadian oats”) was about 60% of the duration of the double-consonant version (“Canadian notes”). This aligns well with previous studies showing that the duration of an ambiguous consonant is one of the primary determinants in whether a particular segment can be found on only one side or on both sides of an acoustically ambiguous prosodic boundary (Fujimura et al., 1978; Repp, 1978; Schouten & Pols, 1983). Double-consonant versions of each of the segmentation-ambiguous versions of each sentence were modified to have a duration about 80% of the originally recorded duration. This was midway between the typical single-consonant and double-consonant versions; the presumption was that such items would be most likely to be potentially ambiguous. The double-consonant versions were chosen to be the baseline because the single-consonant versions, often including vowel-initial words (e.g., *oats*), were more likely to contain additional cues to the location of the word boundary. Vowel-initial words often show irregular phonation and other acoustic signatures of their vowel-initial nature, thus biasing the listener against a double-consonant percept (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996).

Analogous manipulations were performed on segmentally ambiguous versions of each sentence. Word-final voicing was somewhat challenging to make ambiguous. Word-final voiced segments served as the basis, as adding the perception of voicing to word-final voiceless obstruents proved impossible. The duration of the critical vowel in the voiceless-consonant version of word-final segmentally ambiguous sentences (such as “beet knowledge”) was about 82% of the duration of the critical vowel in the voiced-consonant version of the same item (“bead knowledge”). Pilot testing indicated that listeners often did not hear the voiced tokens as ambiguous even with a duration set to be intermediate between the voiced and voiceless tokens (say, 91%). As such, the duration of the critical vowel for word-final voiced tokens was modified to equal about 82% of its originally recorded duration. As these word-final stops were voiced, they often included voicing between the end of the vowel and the stop release when it was present. This voicing interval was replaced with silence, replicating the methods of previous experiments examining word-final stop voicing (e.g., Hillenbrand & Ingrisano, 1984).

For word-initial stops, meanwhile, manipulation was on an absolute scale and began with word-initial voiceless tokens. The VOT for word-initial segment tokens was adjusted to be 20 ms for stops with a VOT of less than 65 ms, 25 ms for stops with a VOT between 65 ms and 80 ms, and 30 ms for stops with all other VOTs. These values were chosen to approximate points of maximal ambiguity in previous studies. The discrete groupings were chosen to very roughly compensate for cue-trading behavior in voicing continua (Repp, 1982).

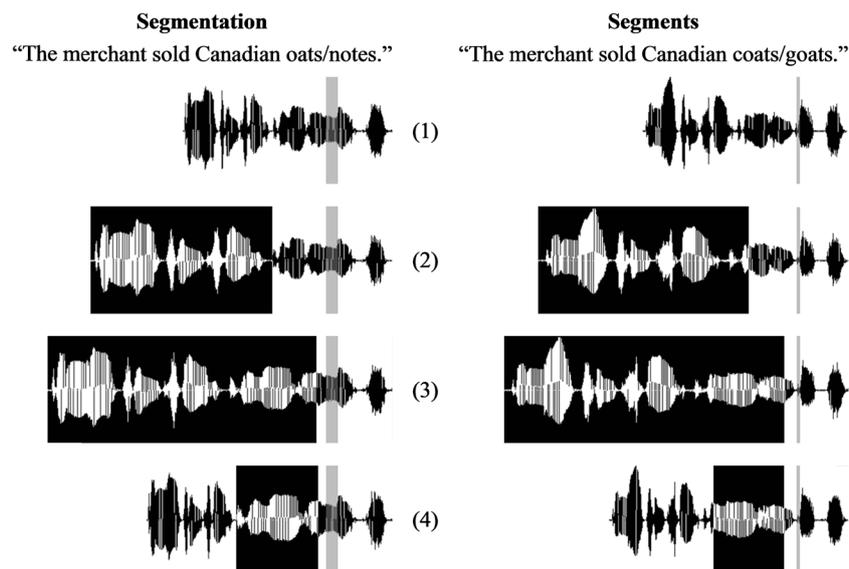
The reasoning was that tokens produced with the longest VOTs should also be maximally ambiguous with a VOT long relative to tokens with the shortest VOTs. For example, in English, /p/ tends to be produced with a shorter VOT than /k/; as such, the point of maximal ambiguity for /p/ segments is at a lower VOT than that for /k/ (Lisker & Abramson, 1967).

From these modified tokens, multiple versions of each token were created that varied in their speech rate, depicted below (Fig. 1). The following experiments differed in which portions of the context were modified; each experiment has a materials section that enumerates the context definition used in that experiment. Three separate versions of the sentence were created: one with the context duration played at 100% of original duration (normal), another at 150% of original duration (slow), and another at 200% of original duration (slower). These stimuli were created using the Pitch-Synchronous Overlap and Add (PSOLA) technique in Praat (Boersma & Weenink, 2009). Filler items, meanwhile, were uniformly rate modified at each of the three possible duration levels. Items were eliminated from the analysis if they had no variation in participant responses within each combination of context definition and ambiguity type; these items had *no* ambiguity with regard to word segmentation. Discarding these items was also necessary for model convergence. This resulted in the elimination of clusters from analysis for each combination of distal definition and ambiguity type, which are listed for each experiment in the Appendix.

## Design and procedure

This study had one between-subjects variable and three within-subjects variables, leading to a 2 (participant group: UMD or MTurk)  $\times$  2 (ambiguity type: segment or segmentation)  $\times$  3 (distal rate: normal, slow, slower)  $\times$  2 (position: word-initial or word-final) mixed design. Participant group varied on a between-participant basis, as participants performed their task either through Mechanical Turk or in the lab, never both. Participants were recruited separately for each group. This opens up some opportunities for differences between the participant groups, as the groups were recruited in a different manner. Ambiguity type, position, and distal rate all varied on a within-participants basis: participants heard every possible combination of ambiguity type, distal rate, and position. Each participant was assigned to one of six possible lists that counterbalanced items across combinations of ambiguity type and distal rate; participants were approximately evenly distributed across lists (the number of participants per list ranged from four to seven).

All participants heard 60 experimental items and 60 filler items. The filler items were evenly distributed across possible values for duration: 100% of original duration (*normal*), 150% (*slow*), and 200% (*slower*). The experimental items were evenly distributed across every possible combination



**Fig. 1** Waveforms for the “Canadian oats/notes/coats/goats” sentence across different definitions of distal context, speech rates, and ambiguity types. In each waveform, the ambiguous portion of the utterance is shown in gray; the rate-modified distal portion of the sentence is inverted. Waveforms on the left show the word segmentation ambiguity condition; waveforms on the right show the segment ambiguity condition. From top to bottom for each ambiguity type condition, the figure shows (1) the unmodified condition, used in all four experiments reported here

(i.e., 100% distal context rate by any definition), (2) the slowest distal context rate for the definition used in Experiment 1, (3) the slowest distal context rate for the definition used in Experiment 2, and (4) the slowest distal context rate for the definition used in Experiment 3. Slight differences in amplitude across context versions reflect that each item was amplitude normalized separately, which led to slightly different overall amplitudes across the files; small differences in timing across each version reflect short ramps created at points of rate changes within the stimuli

of distal rate and ambiguity type, such that there were 10 items in each possible combination. For example, each participant heard 10 items with a slower distal rate and a segmentation ambiguity type. No participant heard more than one version of any cluster. The order of presentation was randomized for each participant across filler and experimental trials. Participants were told to type the last two words that they heard after listening to each sentence and were allowed to repeat the sentences up to 5 times before beginning to write anything down. In person, participants were seated in a quiet room and used Sennheiser M40fs headphones to complete the study. Trials were administered using PsychoPy software (Peirce, 2007). In total, the study usually lasted about 20 to 25 minutes for participants in person.

On Mechanical Turk, the procedure at UMD was replicated as closely as possible within the parameters of Ibex experimental software, written by Alex Drummond (available at <http://spellout.net/ibexfarm/>). Technical limitations prevented certain aspects from being duplicated. Participants were allowed to repeat each sentence fragment as many times as desired, rather than only up to 5 times. Although participants were urged to avoid doing so, it was also possible for them to interrupt the playing of a clip at any point while it was playing, and to start playing the clip at any point in the recording. Further, although listeners were asked to perform the study at a comfortable listening level and to use headphones, listeners were able to set the volume on their headphones or

speakers to whatever level they desired. Some reported not using headphones despite the direct and repeated request. Although the studies generally lasted about 30 minutes for Mechanical Turk participants, some took up to 55 minutes; it is possible that listeners paced their participation by taking breaks or engaging in secondary tasks.

## Analysis

For each trial, an index of the accuracy of each trial was computed. Trials were characterized as “accurate” if the lexical material actually written down had only minimal differences from the intended lexical material. Differences were defined in terms of the features typically used to describe phonetic segments: voicing, place of articulation, and manner of articulation for consonants and height, backness, roundness, and tenseness for vowels. If any segments within a transcription differed from the originally recorded segments in more than one feature, if the transcription included more than one phoneme insertion or deletion, or if the transcription combined any of these possible changes, that trial was discarded from further analysis. For example, transcribing “warp path” for the last two words of the phrase “The travelers enjoyed the wharf path/bath,” was counted as accurate, because the [p] at the end of “warp” differs from the [f] at the end of “wharf” by only the manner of articulation. However, something like “wars path,” which involves a change in place and voicing

in the final consonant of the first word, and “warped path,” which involved both a substitution and an insertion, were excluded. Transcriptions that consisted of only a portion of the intended critical phrase were still included if the region surrounding the critical segment was intact; transcribing “Neapolis sale” for an item intended to be “Minneapolis sale” did not automatically lead to a trial being thrown out.

Accurate tokens were scored with an index that will be called the longer response proportion. This denotes the lexical information actually perceived by the participant. The interpretation of a “longer response” was intended to be consistent across the different conditions: a participant’s transcription was coded as 1 if it indicated they heard some part of the critical segment as relatively long compared to the context, while the transcription was coded as 0 if it was perceived to be relatively short compared to the context. For segmentation contrasts, a relatively long segment was one that could be perceived as straddling a word boundary. As such, a value of 1 was assigned to a sentence if participants transcribed two consonants on either side of a word boundary (e.g., “Canadian notes”), while a value of 0 was assigned if participants transcribed a single consonant on only one side of the word boundary (“Canadian oats”). For word-initial segment contrasts, consonants perceived as having a relatively long VOT should be identified as voiceless. Tokens with voiceless initial consonants (“Canadian coats”) were assigned a value of 1, while tokens with voiced initial consonants (“Canadian goats”) were assigned a value of 0. Finally, for word-final segment contrasts, tokens should be identified as voiced when they are preceded by relatively long vowels. Tokens with voiced final consonants (“bead knowledge”) were assigned a value of 1, while tokens with voiceless final consonants (“beet knowledge”) were assigned a value of 0.

Generalized linear mixed-effects models, implemented using the lme4 (Bates, Maechler, Bolker, & Walker, 2014) package in the R programming language (Version 3.2.2), were used for data analysis. Linear mixed-effects models allow for variation in both fixed effects, which are similar to the traditional main effects and interactions used in more traditional statistics, and random effects, which can account for variation by item and by participant simultaneously. This class of models is said to offer many advantages over ANOVAs and other traditional methods (Baayen, Davidson, & Bates, 2008; Quené & van den Bergh, 2008). The first step was to determine the random effects in the model. Some statisticians have suggested the idea that random effect structures should always be *maximal*; that is, all possible random intercepts and slopes should be included within the models under consideration (Barr, Levy, Scheepers, & Tily, 2013). However, we share recent skepticism about the merits of such a rigid recommendation (Bates, Kliegl, Vasishth, & Baayen, 2015). Bates et al. (2015) argued that using maximal models could lead to massive complexity in model structure, complexity not supported

by the data that would be available. Instead, Bates et al. (2015) proposed a series of steps, with a particular emphasis on model comparison, to determine the ideal random effects structure for an experiment. This approach is familiar to those who have used a model-comparison approach for fixed-effects structures.

To analyze the dataset used here, the most complex fixed and random effect structure possible served as a launching point. For the “initial” model, both random slopes for items and participants were included to reflect individual differences between participants and idiosyncratic effects by item in the effects observed here. Random slopes for distal rate both by participant and by cluster and for participant group by cluster were considered in the full model. To determine the ideal random effect structure in this dataset, a Principal Components Analysis (PCA) was performed on the random effects structure using the RePsychLing package in R (Baayen, Bates, Kliegl, & Vasishth, 2015) to give a rough estimate of the ideal dimensionality of the random effects structure in the data. The random effects were cut out that explained the least variation in the initial model, and an ANOVA was used to examine differences in model fit. In cases when there was no significant difference between the initial model and a model with a less complex random effect structure, and where the PCA showed no justification for having a fuller model with so many random slopes, we found no evidence for the initial model over the less full model, which will be referred to as the “intermediate” model. To reduce repetition, the steps that led to the intermediate model are omitted for all but the first reported results, and a summary of the random effects included in each intermediate model is presented within a table at the beginning of the results section for each study.

The intermediate model was then used as a base from which to determine the ideal fixed effect structure of the model. Every intermediate model includes fixed effects of distal rate and participant group as well as the interaction between them. A main effect of distal rate would imply that the manipulation of context speech rate was successful. Distal rate was coded for model comparison in a continuous fashion, as treating it as a level with three factors would obscure the relationship between the normal, slow, and slower rates. In this cases, “normal” was coded as 0.0, “slow” as 0.5, and “slower” as 1.0. A main effect of participant group would suggest that participants on Mechanical Turk were using different strategies from participants at UMD in performing the study. And a fixed interaction between them would indicate that the effects of distal rate depended on the participant group; that is, that Mechanical Turk participants would differ from UMD participants in the strength of their distal rate effects. Again, the significance of each of these effects was determined through a subtractive approach in model comparison. The fixed parameters of the

winning model are also presented in a summary table at the beginning of each results section.

## Experiment 1

The effects of distal speech rate on segments and segmentation were first probed using the definition of distal rate adapted from Newman and Sawusch (1996). We reasoned that using the most conservative definition of distal information in the previous literature—that is, the one with rate effects most removed from the ambiguous ones—would indicate with more confidence the strength of distal rate effects across both types of ambiguity. Although studies of segments had not shown effects of distal rate context on voicing perception using this definition of distal rate, the strength and persistence of distal rate effects in the segmentation literature led us to expect that segmentation effects should be present even under this more conservative definition of distal. Under representational and processing-based accounts of the differences between segment and segmentation studies, it would be predicted that distal effects on segmentation should be present, while effects on segments should be absent. Under an account stemming from position differences, it might be expected that distal effects should be present for segmentation and for word-final segments, but not for word-initial segments. If the differences come from differences in the definition of distal context, meanwhile, it might be the case that this condition should lead to no distal rate effects across segments and segmentation, as this more conservative definition of distal resembles more closely the definitions of distal context that led to null effects in previous studies of segment ambiguities.

## Method

**Participants** Twenty-four participants at the University of Maryland, College Park (UMD) were recruited to participate in this study. Most participants were recruited for course credit; others were compensated \$8 for their participation in this study and another, unrelated speech perception experiment. All participants self-reported normal hearing. Six participants reported not being native monolingual English speakers and were excluded. This left, in total, 18 participants (age  $M = 25.1$ ; range: 18–48; 14 female, four male). Most of the remaining participants (11, 61%) reported their primary state of residence growing up (as defined by the state in which they spent the greatest number of years living in between the ages of 0 and 18 years) as Maryland; no other states were represented by more than two participants. This experiment and all subsequent experiments in this paper were vetted by the University of Maryland, College Park's Institutional Review Board (IRB).

Amazon's Mechanical Turk system was used to recruit 18 participants to participate in this experiment on the Internet. All participants were compensated \$4 for their participation. All participants self-reported normal hearing and native English proficiency, although four were excluded for not scoring sufficiently high on a test to determine their status as English native speakers. The 14 participants remaining were, on average, 28.6 years old (range: 24–37; median: 27.5; seven female, seven male). The top reported location of primary residence was Pennsylvania (4, 28.6%), with no other state exceeding two participants. None came from other countries. According to their self-reports, of the 14 participants who passed the English native speaker test, three used their computer's preinstalled speakers, two used supra-aural headphones, three used circumaural headphones, and six used earbuds.

**Materials** For Experiment 1, a conservative definition of the region of speech that was considered *distal* was adopted, a definition that was similar to that of Newman and Sawusch (1996). Under this conservative definition, the distal context was defined as anything lying before the first syllable onset before a point 400 ms previous to the onset of the critical segment. That is, anything within 400 ms was considered proximal, and the *distal* context was every syllable that was completed prior to that point. For this definition of *distal*, 20 segmentation and 20 segmental items were excluded due to a lack of variation in participant responses, as reported in the Appendix.

## Results

Tables 1 and 2 give random and fixed-model parameters for each combination of ambiguity type and position. These are reported in summary form here to avoid unnecessary repetition; however, the precise analysis stream by which these effects were uncovered is also described in the first results section to ensure the procedure is clear.

**Segmentation** Figure 2 summarizes the average longer response rates by position, participant group, and distal rate.

**Word-initial.** First, the optimal random effects structure given the dataset was probed using the tools suggested in Bates et al. (2015). The initial model included all fixed effects of interest as well as random intercepts by cluster, random slopes for proximal rate and participant group by cluster, random intercepts by participant, and random slopes for proximal rate by participant. A principal components analysis (PCA) was performed on the variance-covariance matrix of the model, which indicated that one dimension was sufficient to explain random variation in the model by participant and that at most two were sufficient by cluster. Indeed, comparing the initial model to one with only random intercepts by cluster

**Table 1** Random parameters included in the intermediate models for different combinations of ambiguity type and position in Experiment 1. Check marks indicate random parameters that were included in the intermediate model, while crosses indicate parameters that were excluded.

Ambiguity	Position	By participant		By cluster		
		Intercepts	Slopes for distal rate	Intercepts	Slopes for distal rate	Slopes for group
Segmentation	Initial	√	×	√	√	×
Segmentation	Final	√	×	√	√	×
Segment	Initial	√	√	√	×	×
Segment	Final	√	×	√	×	×

and by participant and random slopes by cluster for distal rate showed no significant loss of model fit,  $\chi^2(5) = 0.0725$ ,  $p = 1$ . This model was dubbed the “intermediate model.”

The intermediate model was then compared with models lacking each of the fixed effects. Main effects were assessed by comparing the intermediate model with a model that lacked both main effects and interactions with the factor in question. We tested for effects of distal rate by comparing the intermediate model to a model that lacked fixed effects of distal rate and the two-way interaction between distal rate and participant group. Comparing the intermediate model to one without the fixed effect of distal rate (or its interaction with participant group) showed no significant difference in model fit,  $\chi^2(2) = 1.04$ ,  $p = .59$ . Thus, variation in distal rate did not significantly impact participants’ perception of the critical region. There was also no significant difference between this model and one that lacked a fixed effect of participant group,  $\chi^2(1) = 1.44$ ,  $p = .23$ . As such, it is reasonable to conclude that neither distal rate nor participant group had an effect on participants’ likelihood of hearing a “long” percept in the critical region for word-initial segmentation ambiguities using the most conservative definition of *distal* rate.

**Word-final.** As before, there was no evidence for a main effect of distal rate,  $\chi^2(2) = 4.53$ ,  $p = .10$ . Nor was there any evidence for an effect of participant group when compared to the model without distal rate,  $\chi^2(1) = 0.0102$ ,  $p = .92$ . Thus, again, the best fitting model is one with only an intercept.

**Segments** Figure 3 summarizes the average longer response rates by position, participant group, and distal rate.

**Word-initial.** There was no significant decrease in model fit between the intermediate model and a model without simple main effects of distal rate and its interaction with participant group,  $\chi^2(2) = 2.30$ ,  $p = .32$ . Indeed, this model without distal rate did not show a significant improvement in fit over the most sparse model that was considered, with only an intercept,  $\chi^2(1) = 0.186$ ,  $p = .67$ . This implies that neither distal rate nor participant group had a significant effect on the perception of the critical region for these items.

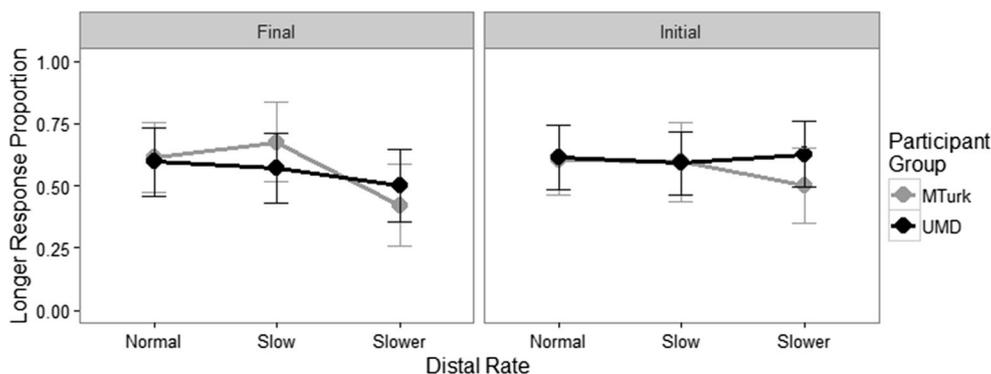
**Word-final.** There was only a marginally significant effect of distal rate on longer percept report rates,  $\chi^2(2) = 4.81$ ,  $p = .09$ , indicating that distal rate may not have significantly influenced participants’ likelihood to hear a longer percept in the target region. However, comparing the intermediate model to one without participant group did yield a significant drop in model fit,  $\chi^2(2) = 6.55$ ,  $p = .04$ . The marginal significance of the distal rate effect led to further exploration of other models to fit the data. The model with the lowest AIC value included both simple effects of participant group and distal rate, but not the interaction between them. This model did not have a significantly worse fit than the intermediate model,  $\chi^2(1) = 0.223$ ,  $p = .64$ , but did have better model fit than both the model lacking the simple effect of distal rate,  $\chi^2(1) = 4.59$ ,  $p = .03$ , and the model lacking the simple effect of participant group,  $\chi^2(1) = 6.32$ ,  $p = .01$ .

## Discussion

Experiment 1 provided surprising results. There was no support for distal effects on segmentation for this conservative definition of *distal*. However, there were such effects present

**Table 2** Fixed parameter estimates for different combinations of ambiguity type and position within Experiment 1

Ambiguity	Position	Intercept			Distal rate			Group: UMD		
		<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>
Segmentation	Initial	0.433	1.10	.27	–	–	–	–	–	–
Segmentation	Final	0.472	0.852	.39	–	–	–	–	–	–
Segment	Initial	0.316	0.993	.32	–	–	–	–	–	–
Segment	Final	0.692	1.37	.17	-1.04	-2.62	.009	-0.787	-2.18	.03



**Fig. 2** Proportion of longer responses for segmentation trials using the most conservative definition of “distal” context by distal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors

for segment items, but only for the perception of word-final segment voicing, not word-initial voicing. Additionally, a main effect of participant group also emerged for word-final segment items. This indicates that distal rate effects *can* be present in segment contrasts; they just emerge only for the word-final voicing contrasts that have not previously been a frequent topic of discussion. These facts argue against a representational or processing account for the differences in effect sizes, as the effects here were present for segments, but not segmentation. They also provide evidence that suggests the importance of position on distal rate effects.

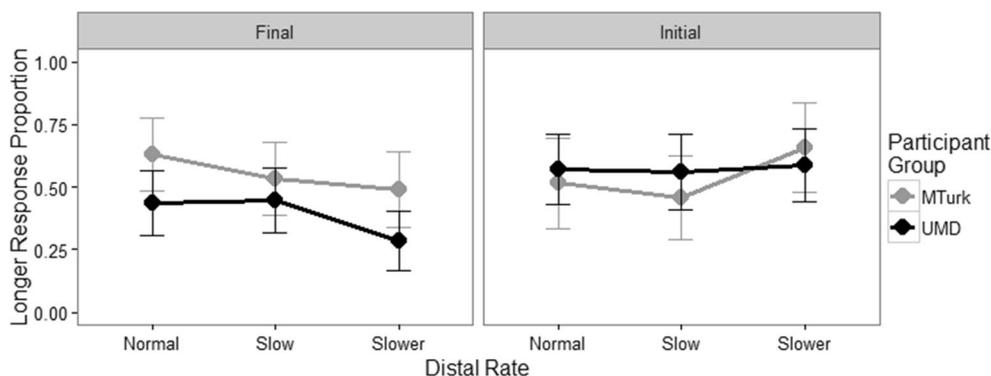
### Experiment 2

The surprising lack of significance in distal rate effects for segmentation items observed in Experiment 1 suggested that at least some of the differences in rate effects between previous studies of segments and segmentation may have been the result of differences in the definition of the distal context. Experiment 1, which incorporated the definition of distal used in a study of segments, largely replicated the results of previous studies of segments (i.e., null effects). In Experiment 2, a more liberal definition of distal context with regard to the amount of context considered distal was adopted, taken from

the definition used in Dilley and Pitt (2010). This definition entailed modifying the speech rate of information more than one syllable removed from the segmentation ambiguity in question. Finding a distal rate effect here where one was not found in Experiment 1 would indicate that at least part of the differences between studies of segments and segmentation are the result of the definition of distal context. That is, the finding of “distal” effects in prior segmentation studies might not be from truly distal information. In particular, it was expected based on previous studies that segmentation items should show distal rate effects in a way that should not be present for at least word-initial segment items.

### Method

**Participants** Similar samples were taken to those used in Experiment 1. Twenty-eight participants were recruited to participate in this study at the University of Maryland, College Park (UMD). Most participants were recruited for course credit; others were compensated \$8 for their participation in this study and another, unrelated speech perception experiment. All participants self-reported normal hearing. Five participants reported not being native monolingual English speakers and were excluded; three were excluded due to experimenter or participant error. This left, in total, 20 participants (age  $M =$



**Fig. 3** Proportion of longer responses for segment trials using most conservative definition of “distal” context by distal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors.

21.0; range: 19–33; nine female, 10 male, one not stated). Most of the remaining participants (15, 75%) reported their primary state of residence growing up as Maryland; no other locations were represented by more than one participant.

Amazon's Mechanical Turk system allows for the recruitment of 18 participants to participate in this experiment on the Internet. All participants were compensated \$4 for their participation. All participants self-reported normal hearing and native English proficiency, although one was excluded for not scoring sufficiently high on a test to determine their status as English native speakers. The 17 participants remaining were, on average, 37.0 years old (range: 22–67; median: 32; seven female, 10 male). Top reported locations of primary residence included California (5, 29.4%), New York (5, 29.4%), and Texas (4, 23.5%). None came from other countries. Participants were asked to wear headphones when participating in the experiment; according to their self-reports, of the 17 participants who passed the English native speaker test, two used their computer's preinstalled speakers, three used supra-aural headphones, six used circumaural headphones, and six used earbuds.

**Materials** In Experiment 2, the liberal definition of distal with regard to the amount of context labeled as such, which was used by Dilley and Pitt (2010), was adopted, with the distal context defined as anything more than one syllable previous to the onset of the critical segment. Seventeen segmentation items and 18 segment items were excluded for the present distal definition, given in the Appendix.

## Results

Tables 3 and 4 give the random and fixed parameters for the best fitting models in Experiment 2. The process of finding these models, and figures illustrating trends in each experiment, are described in more detail below.

**Segmentation** Figure 4 summarizes the average longer response rates by position, participant group, and distal rate for items with segmentation ambiguity.

**Word-initial.** Comparing the intermediate model for word-initial segmentation ambiguities to a model without distal rate or its interaction with participant group showed a significant decrease in model fit as a result of the removal of the distal rate fixed effect,  $\chi^2(2) = 34.9, p < .001$ . However, the comparison of the intermediate model with a model lacking effects of participant group did not show a significant effect,  $\chi^2(2) = 0.0145, p = .99$ . This suggests that distal rate, but not participant group, had a significant influence on how likely participants were to report “longer” percepts within the critical region.

**Word-final.** Comparing the intermediate model for word-final segmentation ambiguities to one without any fixed distal

rate effects showed a significant decrease in model fit after distal rate was removed from the model,  $\chi^2(2) = 12.3, p = .002$ . The same cannot be said for a model removing the fixed effects of participant group, which performed no differently from the intermediate model,  $\chi^2(2) = 0.191, p = .91$ . As such, the best fitting model here only included fixed effects of distal rate.

**Segments** Figure 5 summarizes the average longer response rates by position, participant group, and distal rate for items with segment ambiguity.

**Word-initial.** The intermediate model did not fit the data for word-initial segment contrasts better than a model that lacked the fixed effect of distal rate and the interaction between distal rate and participant group,  $\chi^2(2) = 3.13, p = .21$ . Nor did this simpler model explain more variation than did a model without an effect of participant group,  $\chi^2(1) = 1.21, p = .27$ . Therefore, it appears that there is not significant evidence for either distal rate or participant group affecting participants' perception of word-initial segments, even under a relatively liberal definition of distal rate.

**Word-final.** The effects of distal rate were only marginal,  $\chi^2(2) = 4.82, p = .09$ . There was no evidence for effects of participant group with a model with just an intercept explaining no less variation than the intermediate model,  $\chi^2(2) = 3.92, p = .14$ . The marginal significance of the difference between the intermediate model and the one lacking distal rate led to a decision to construct models with every combination of each fixed factor. This time, however, the model with the lowest AIC value was the one with only an intercept. This model did not perform significantly worse than the intermediate model,  $\chi^2(3) = 4.86, p = .18$ , nor than the model with the second-lowest AIC value, which had only a simple effect of distal rate,  $\chi^2(1) = 0.942, p = .33$ .

## Discussion

Experiment 2 shows effects more familiar than those of Experiment 1: Segmentation was affected by distal rate, while segments were not (both word-initially and word-finally). The segmentation effects are well in line with previous studies of segmentation in the literature (Dilley & Pitt, 2010), as is the failure to find effects for word-initial voicing contrasts. Experiment 1 and Experiment 2 therefore suggest that information very distal from a point of ambiguity does not typically (or powerfully) affect the perception of either segmentation or segments, whereas somewhat-distal information seems to affect segmentation alone. If this holds, it might suggest that segmentation decisions rely on a larger processing window than do segmental ones, but that neither are typically impacted (or, at least, impacted strongly) by durational information far from the point of ambiguity.

**Table 3** Random parameters included in the intermediate models for different combinations of ambiguity type and position in Experiment 2. Check marks indicate random parameters that were included in the intermediate model, while crosses indicate parameters that were excluded.

Ambiguity	Position	By participant		By cluster		
		Intercepts	Slopes for distal rate	Intercepts	Slopes for distal rate	Slopes for group
Segmentation	Initial	√	×	√	×	×
Segmentation	Final	√	×	√	×	×
Segment	Initial	√	√	√	√	√
Segment	Final	√	×	√	√	×

That made a third distal rate experiment even more important.

### Experiment 3

Experiment 2 differed from Experiment 1 in distal rate effect sizes for both segmentation (absent in Experiment 1, present in Experiment 2) and word-final segments (present in Experiment 1, absent in Experiment 2). This indicates that at least part of the differences between segmentation and segment studies in the prior literature result from the definitions of “distal context” used by each type of experiment rather than because of other properties of the stimuli or of the processing systems more generally. The differences between Experiment 1 and Experiment 2 warrant an explanation: what information was available for listeners in Experiment 2 that was unavailable in Experiment 1 for segmentation, or vice versa for word-final segments? One explanation is that it is differences in the *proximity* between the distal context definitions used in Experiment 1 and Experiment 2. That is, the distal rate effects in Experiment 2 were stronger because the definition of distal rate included information that was closer in time to the ambiguities in question. If so, it might be expected that this “nearby” distal information would be sufficient to cause effects by itself.

Alternatively, it may be the case that it is differences in the *consistency* of the distal context that triggered the differences between Experiment 1 and Experiment 2. In Experiment 2, the context rate was relatively consistent; a wide swath of the distal context was present at a single, unchanging rate. However, in Experiment 1, the rate of speech of the context

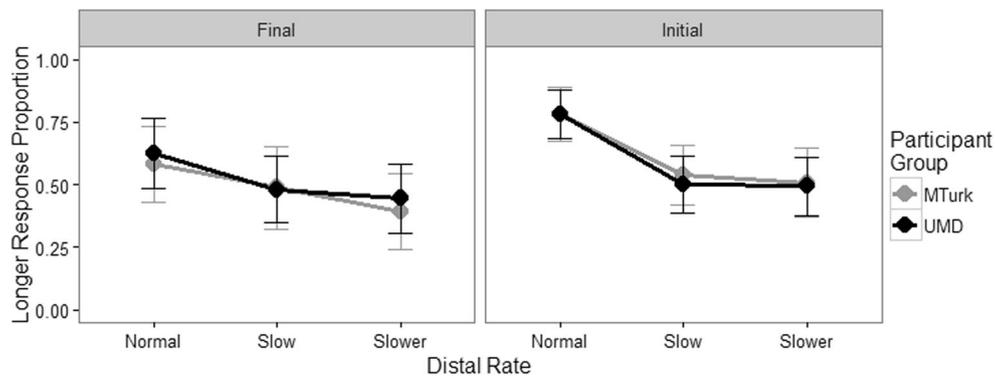
went from slow to unmodified at an earlier point in the sentence, meaning that the rate of the distal context was inconsistent across the sentence. It might be that this inconsistency, not the degree of locality to the critically ambiguous portions of each item, led to the differences in distal rate effects between Experiment 1 and Experiment 2. As such, in Experiment 3, a distal context was designed that reflected the difference in distal context definitions between Experiment 1 and Experiment 2. This allowed for the differentiation of the effects of proximity and consistency in the definition of distal context on segmentation and segments, as this region is closer in time to the critical region than the distal context of Experiment 1 while also being inconsistent with the previously-presented context.

### Method

**Participants** 21 participants were recruited at the University of Maryland, College Park (UMD) to participate in this study. Most participants were recruited for course credit, while others were compensated \$8 for their participation in this study and another, unrelated speech perception experiment. All participants self-reported normal hearing. One participant reported not being a native monolingual English speaker and was excluded. This left, in total, 20 participants (age  $M = 22.4$ ; range: 18–36; 13 female, seven male). Most of the remaining participants (13, 65%) reported their primary state of residence growing up (as defined by the state in which they spent the greatest number of years living in between the ages of 0 and 18) as Maryland; no other states were represented by more than two participants. One participant reported growing up in

**Table 4** Fixed parameter estimates for different combinations of ambiguity type and position within Experiment 2

Ambiguity	Position	Intercept			Distal rate			Group: UMD		
		<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>
Segmentation	Initial	1.47	3.67	< .001	-1.90	-5.53	< .001	—	—	—
Segmentation	Final	0.593	1.47	.14	-1.21	-3.43	< .001	—	—	—
Segment	Initial	-0.110	-0.373	.71	—	—	—	—	—	—
Segment	Final	-0.0390	-0.114	.91	—	—	—	—	—	—



**Fig. 4** Proportion of longer responses for segmentation trials using the most liberal definition of “distal” context by distal rate (*horizontal axis*), positions (*columns*), and participant group (*shade*). Error bars show by-participant standard errors

Jamaica. One participant was excluded specifically for word-initial segmentation items because that participant was completely at ceiling in terms of the longer response proportion rate and thus prevented model convergence.

Amazon’s Mechanical Turk system was used to recruit 19 participants to participate in this experiment on the Internet. All participants were compensated \$4 for their participation. All participants self-reported normal hearing and native English proficiency, although three were excluded for not scoring sufficiently high on a test to determine their status as English native speakers. The 16 participants remaining were, on average, 34.6 years old (range: 22–54; median: 35; four female, 12 male). The top reported location of primary residence was Tennessee (4, 25%), with no other state exceeding two participants. None came from other countries. Participants were asked to wear headphones when participating in the experiment; according to their self-reports, of the 16 participants who passed the English native speaker test, one used their computer’s preinstalled speakers, two used external speakers, five used supra-aural headphones, three used circumaural headphones, and four used earbuds.

**Materials** In Experiment 3, the duration of the portion of the sentence that fell between the more conservative definition of distal used in Experiment 1 and the more liberal definition of

distal used in Experiment 2 (i.e., the portion more than one syllable before the critical phoneme, but less than 400 ms from it) was manipulated to examine whether that region alone could explain any effects obtained under the liberal definition. A total of 19 segmentation items and 14 segment items were removed from Experiment 3. They are listed in the Appendix.

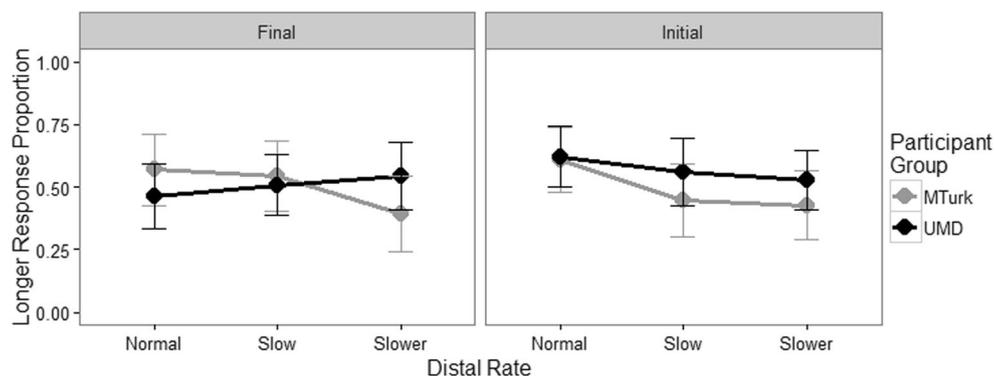
## Results

Tables 5 and 6 give the random and fixed parameters for the best fitting models in Experiment 3.

**Segmentation** Figure 6 summarizes the average longer response rates by position, participant group, and distal rate.

**Word-initial.** Comparing the intermediate model for word-initial segmentation ambiguities to one lacking effects of distal rate led to a significant decrease in model fit,  $\chi^2(2) = 17.4, p < .001$ . However, the intermediate model fit no better than one lacking fixed effects of participant group,  $\chi^2(2) = 0.431, p = .81$ . This suggests that distal rate made a significant impact on the perception of the critical region, but that participant group did not.

**Word-final.** Removing fixed distal rate effects from the intermediate model for word-final segmentation ambiguities



**Fig. 5** Proportion of longer responses for segment trials using the most liberal definition of “distal” context by distal rate (*horizontal axis*), positions (*columns*), and participant group (*shade*). Error bars show by-participant standard errors

**Table 5** Random parameters included in the intermediate models for different combinations of ambiguity type and position in Experiment 3. Check marks indicate random parameters that were included in the intermediate model, while crosses indicate parameters that were excluded.

Ambiguity	Position	By participant		By cluster		
		Intercepts	Slopes for distal rate	Intercepts	Slopes for distal rate	Slopes for group
Segmentation	Initial	√	×	√	×	×
Segmentation	Final	√	×	√	×	×
Segment	Initial	√	×	√	√	×
Segment	Final	√	√	√	×	×

significantly hurt model fit,  $\chi^2(2) = 18.7, p < .001$ . The same was not true for participant group, which did not significantly impact model fit,  $\chi^2(2) = 1.19, p = .55$ . As such, it is reasonable to conclude that distal rate, but not participant group, affected the likelihood that people gave a long report of the critical region for word-final segmentation contrasts.

**Segments** Figure 7 summarizes the average longer response rates by position, participant group, and distal rate.

**Word-initial.** Comparing the intermediate model for word-initial segment ambiguities to one lacking fixed effects of distal rate showed no significant difference,  $\chi^2(2) = 3.50, p = .17$ . There was no difference between that model and one that included only an intercept,  $\chi^2(1) = 0.584, p = .44$ . Therefore, the model with only the interaction appears to be the best fitting one, suggesting that neither distal rate nor participant group influence the perception of word-initial segments for these stimuli.

**Word-final.** There was no significant difference between the intermediate model for word-final segment ambiguities and a model lacking fixed effects of distal rate,  $\chi^2(2) = 4.20, p = .12$ . And there was only a marginally significant difference between that model and one lacking fixed effects of participant group, a model with just an intercept,  $\chi^2(1) = 3.27, p = .07$ . Again, the marginal significance of many of these effects led to the decision to try a multitude of different models to fit this dataset. In this case, the model with the smallest AIC value was a model with simple effects of participant group and distal rate but no interaction between them. This model fit the data no worse than the intermediate model,  $\chi^2(1) = 0.107, p = .74$ , but significantly better than the model lacking

the simple effect of distal rate,  $\chi^2(1) = 4.09, p = .04$ , and marginally better than the model lacking the simple effect of participant group,  $\chi^2(1) = 2.90, p = .09$ . As such, with some hesitation, the tentative conclusion is that this model (the intermediate model without the distal rate/participant group interaction) provides the best fit to the dataset.

**Discussion**

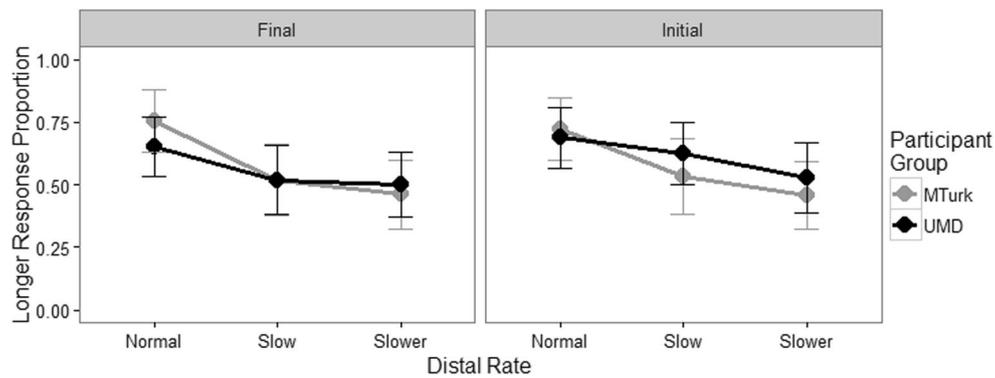
The results of Experiment 3 represent an amalgamation of the results of Experiments 1 and 2. Like in Experiments 1 and 2, there were no effects of distal rate on word-initial segment contrasts. Like in Experiment 2, there were distal rate effects on segmentation. And, like in Experiment 1, there were distal rate effects on word-final segments. For segmentation, this suggests that it is the “near-distal” region—close enough to be considered distal for previous studies of segmentation but not far enough away to be considered distal for previous studies of segments—that plays the primary role in determining critical region perception. Variation in the duration of this section alone was sufficient to cause an effect on listeners’ judgments. For segments, it appears that the effects of distal rate are weaker, present only for word-final contrasts, and only sporadic in nature.

**Experiment 4**

The previous experiments lack a direct comparison between segments and segmentation for the distal context definitions adopted. It would certainly be of interest if differences in effect sizes between segments and segmentation gave

**Table 6** Fixed parameter estimates for different combinations of ambiguity type and position within Experiment 3

Ambiguity	Position	Intercept			Distal rate			Group: UMD		
		<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>
Segmentation	Initial	1.20	2.86	.004	-1.52	-3.98	< .001	–	–	–
Segmentation	Final	1.02	2.36	.02	-1.46	-4.10	< .001	–	–	–
Segment	Initial	0.227	0.725	.47	–	–	–	–	–	–
Segment	Final	0.557	1.47	.14	-0.715	-2.16	.03	-0.537	-1.74	.08



**Fig. 6** Proportion of longer responses for segmentation trials using the difference definition of context by distal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors

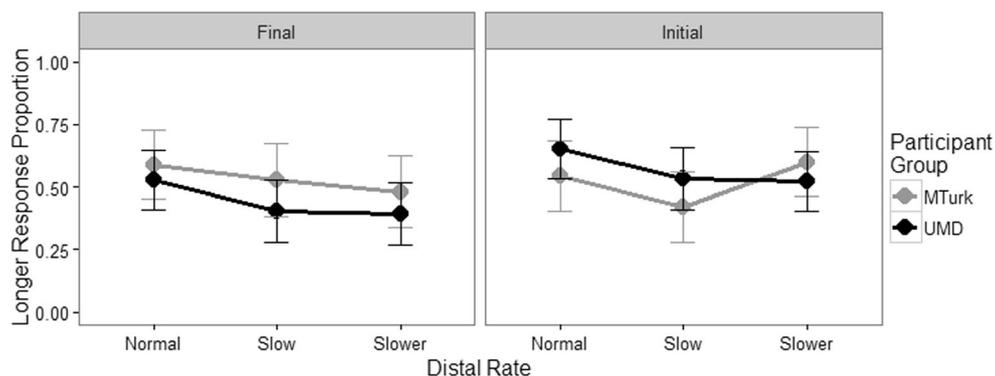
information about differences in representation between individual segment identities and word segmentation. However, it may just be that any uncovered differences resulted from differences in the strength of phonetic cues to segmentation and segments other than the key ones evaluated here, such as proximal acoustic cues. These cues may trade off differently with the cue of distal speech rate, in the style of previous phonetic cue trading experiments (Miller, 1994; Repp, 1982). To assess the ease with which the segment and segmentation items could be directly compared, a different manipulation was employed that we believed would influence segmentation and segments equally: proximal duration. In particular, the duration of the vowel following the critical segment for these sentences was changed. As reviewed in the introduction, proximal context effects are very well-attested, particularly for segments (Miller & Dexter, 1988; Summerfield, 1981; Volaitis & Miller, 1992), but also for segmentation (Dilley & Pitt, 2010; Reinisch et al., 2011). We expected proximal rate effects to be both present and comparable between segmentation and segment items.

## Method

**Participants** Twenty-six participants at the University of Maryland, College Park, were recruited for \$10 compensation

and were also run in an unrelated study about lexical tone learning. All participants self-reported normal hearing and being a native speaker of English. One participant was excluded for prior experience in similar experiments and one participant was excluded because of a technical error, leaving 24 total participants (age  $M = 22.6$ ; range: 18–28, with one participant aged 45; 18 female, six male). Again, these participants were largely representative of the typical UMD student body, with 19 (79%) reporting their primary residence growing up being Maryland. No other states were represented by more than two participants.

Amazon's Mechanical Turk system was used to recruit 14 participants to participate in this experiment on the Internet. All participants were compensated \$4 for their participation. All participants self-reported normal hearing and native English proficiency. The 14 participants remaining were, on average, 33.6 years old (range: 25–45; median: 34.5; six female, eight male). Participants' home states, as determined by the primary state of residence before the age of 18, were widely distributed, with the only states having more than one representative being Arizona (2), California (2), and Florida (2). None came from other countries. Participants were asked to wear headphones when participating in the experiment; according to their self-reports, one used a computer's preinstalled speakers, two used supra-aural headphones, five used circumaural headphones, and six used earbuds.



**Fig. 7** Proportion of longer responses for segment trials using the difference definition of distal context by distal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors

**Table 7** Random parameters included in the intermediate models for different combinations of ambiguity type and position in Experiment 4. Check marks indicate random parameters that were included in the intermediate model, while crosses indicate parameters that were excluded.

Ambiguity	Position	By participant		By luster		
		Intercepts	Slopes for distal rate	Intercepts	Slopes for distal rate	Slopes for group
Segmentation	Initial	√	×	√	×	√
Segmentation	Final	√	√	√	√	×
Segment	Initial	√	×	√	√	×
Segment	Final	√	×	√	×	√

**Materials** For Experiment 4, proximal context was manipulated. In this case, proximal was defined as the duration of the next vowel after the critical segment. For example, for the “Canadian oats/notes/coats/goats” cluster, the duration of the [o] vowel in “oats/notes/coats/goats” was manipulated, while in the “bee/bean/beat/bead knowledge” cluster, the duration of the [a] vowel in “knowledge” was changed. The second scenario, for word-final tokens, may seem somewhat counter-intuitive, as the vowel is not actually located in the word that perceptually alternates; however, the definition that we adopted allowed us to keep a constant definition of the proximal context across segmentation and segmental contexts. Unfortunately, this desire for consistency also prevented us from adopting a definition of proximal context that was prior to the ambiguity within the critical region, as many potential definitions of “proximal” involving prior context shared quite a bit of overlap with the definition of distal adopted in Experiment 3. For example, for the cluster *Bailey has much bee/bean/beat/bead knowledge*, a proximal definition of “the immediately preceding syllable” would sometimes include the word *much*, which also fell under the definition of distal context used in Experiment 3. Twenty-three clusters were excluded for the segmentation items, while 16 clusters were excluded for the segment items. The materials excluded are listed in the Appendix.

**Results**

Tables 7 and 8 give the random and fixed parameters for the best fitting models in Experiment 4.

**Segmentation** Figure 8 summarizes the average longer response rates by position, participant group, and distal rate for segmentation ambiguities.

**Word-initial.** Comparing the intermediate model to one without effects of proximal rate yielded no difference in critical region report rates,  $\chi^2(2) = 1.42, p = .49$ . Nor was this model in turn any better at fitting the data than one with only an intercept,  $\chi^2(1) = 0.0073, p = 1$ . Thus, the best fitting model for this data involved only a fixed intercept.

**Word-final.** The intermediate model did no better of a job fitting the data than a model that lacked any influence of proximal rate,  $\chi^2(2) = 4.56, p = .10$ . Nor was the model without effects of distal rate any better at fitting the data than a model with only an intercept,  $\chi^2(1) = 1.48, p = .22$ . Therefore, the best-fitting model for this dataset included only a fixed intercept.

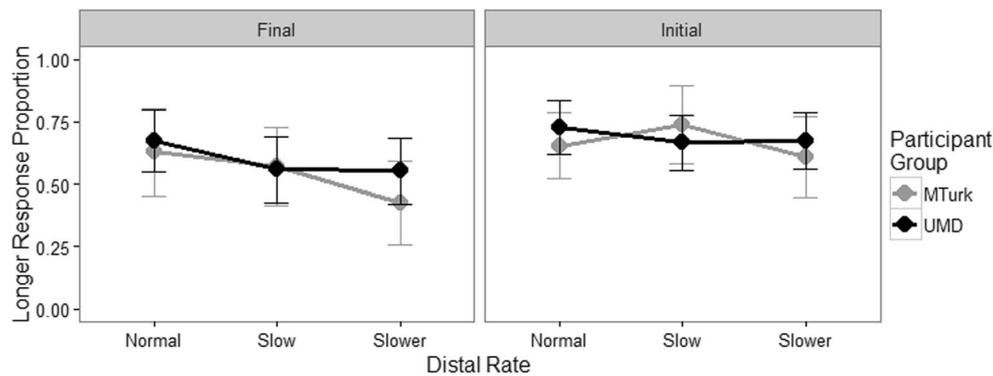
**Segments** Figure 9 summarizes the average longer response rates by position, participant group, and distal rate for segment ambiguities.

**Word-initial.** For fixed effects, the model with no effects of proximal rate did not significantly differ from the intermediate model,  $\chi^2(2) = 1.74, p = .42$ . Nor did this reduced model differ from a model with only an intercept,  $\chi^2(1) = 0.36, p = .55$ . Thus, the best fitting model appears to be the simplest possible one.

**Word-final.** A reduced model with no fixed effects of proximal rate,  $\chi^2(2) = 0.802, p = .67$ , was not a significantly worse fit than the intermediate model, while a minimal model with just an intercept was in turn not a better fit than the reduced

**Table 8** Fixed parameter estimates for different combinations of ambiguity type and position within Experiment 4

Ambiguity	Position	Intercept			Distal rate			Group: UMD		
		<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>	<i>b</i>	<i>z</i>	<i>p</i>
Segmentation	Initial	0.935	2.56	.01	–	–	–	–	–	–
Segmentation	Final	0.491	1.01	.31	–	–	–	–	–	–
Segment	Initial	0.227	-0.0808	-0.192	–	–	–	–	–	–
Segment	Final	0.154	0.379	.71	---	---	---	---	---	---



**Fig. 8** Proportion of longer responses for segmentation trials using proximal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors

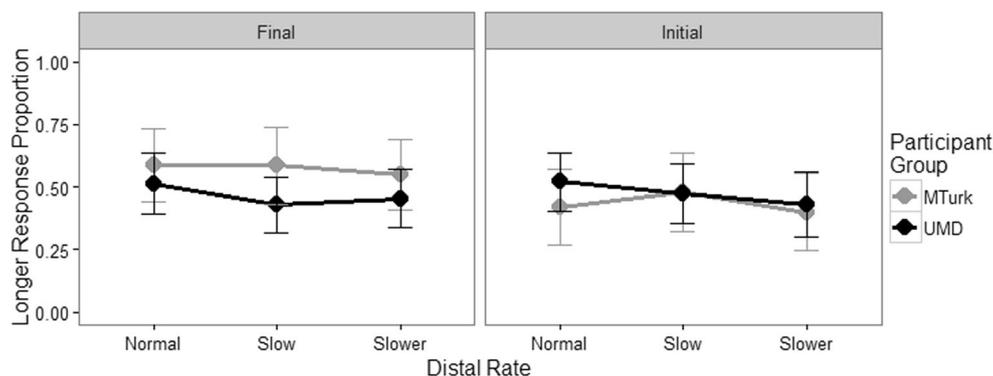
model,  $\chi^2(1) = 2.40, p = .12$ . Thus, the best fitting model was one that just included a fixed intercept.

## Discussion

The materials used in this experiment were not subject to proximal rate effects. This fact is surprising. Proximal rate effects are ubiquitous in the literature (Dilley & Pitt, 2010; Heffner, Dilley, McAuley, & Pitt, 2013; Miller & Dexter, 1988; Newman & Sawusch, 1996; Reinisch et al., 2011; Seidl, 2007; Summerfield, 1981; Toscano & McMurray, 2014; Volaitis & Miller, 1992). The reasons for the failure to find these effects are likely many, and are outside the scope of this article. They may include, for example, the diversity of the sentences used in this experiment or the lack of repetition of individual items. Probing such explanations would require additional experiments. More immediately, this fact also prevents direct comparisons between segment and segmentation items in the analyses used in this paper, as it could not be established conclusively whether the items were directly comparable, as might be suggested by similar proximal rate effects across ambiguity types.

## General discussion

The primary objective of this study was simple: to determine the reasons for the difference between segments and segmentation in the strength of distal rate effects. Four possible explanations were considered. First, that the representation of segments and segmentation led to the differences in distal rate effects; perhaps distal rate effects were stronger for segmentation over segments because the distal rate information was represented in a way that made it “less distal” for segmentation over segments. Second, that it had something to do with the processing of segmentation and segments; perhaps listeners remain more uncommitted for longer for segmentation. A third explanation was that the difference in effect sizes stems from the differences in the definition of distal context between studies of segments (which generally use relatively conservative definitions of “distal” that include only far-away information) and studies of segmentation (which generally use fairly liberal definitions of “distal” and sometimes include more-intermediate information). Fourth and finally, one idea was that the difference in the effects in each case relate more to the types of consonants being manipulated in the segment case; studies of segments generally focus on word-initial stop



**Fig. 9** Proportion of longer responses for segment trials using proximal rate (horizontal axis), positions (columns), and participant group (shade). Error bars show by-participant standard errors

**Table 9** Z scores for the distal rate variable across each combination of experiment, critical consonant position, and ambiguity type (*ns* = nonsignificant)

Exp.	Definition	Position	Segmentation	Segments
1	Conservative	Initial	<i>ns</i>	<i>ns</i>
2	Liberal	Initial	-5.56	<i>ns</i>
3	Difference	Initial	-3.98	<i>ns</i>
1	Conservative	Final	<i>ns</i>	-2.62
2	Liberal	Final	-3.43	<i>ns</i>
3	Difference	Final	-4.10	-2.16

consonants, which might be less subject to influences beyond VOT and other immediately adjacent cues.

It was intended to compare items with segment ambiguities to items with segmentation ambiguities directly. To do this, a stimulus set was created with paired sentences that differed from each other minimally other than in the type of ambiguity (segmentation or segment) that was present. Experiment 4, which was suggested by reviewers as a way to assess if such a comparison could be valid, in fact turned up no support for proximal effects in these materials at all. This is a puzzling finding, as there is abundant evidence for proximal duration effects on both segments (Newman & Sawusch, 1996; Summerfield, 1981; Toscano & McMurray, 2014) and segmentation (Heffner et al., 2013; Seidl, 2007). The reasons for this failure to find an effect are unclear and are likely outside the scope of this investigation, but this failure makes it challenging to conclude that any significant differences within a single experiment between segments and segmentation are not the results of the properties of the individual recordings of each segment and segmentation item. This suggests that the focus in the present discussion should be on comparison of the effects observed across experiments within segmentation items and within segment items for both word-initial and word-final items. Table 9 summarizes the results for each combination of experiment, position, and ambiguity type.

The pattern for the studies of segmentation was fairly straightforward: it appears that listeners privilege short-lag time information in determining word segmentation. In Experiment 1, no significant effect of distal speech rate on the perception of the critical region emerged across segmentation ambiguities using a distal context definition mimicking those previously proposed in the segmental literature (e.g., Newman & Sawusch, 1996). That is, with the distal context defined in terms of a strict 400-ms time window (rounded up to the nearest syllable onset), distal context did not influence segmentation ambiguities. However, Experiment 2, which used a broader definition of the distal context as “anything further than one syllable removed from a point of ambiguity,” showed significant effects of distal context. Experiment 3 confirmed that the effects observed in

Experiment 2 likely resulted in large part from the difference in context definitions between the first two studies: the proximity of context information being modified to the ambiguity being measured matters, while the consistency of the potentially distal context does not. This follows the results of Reinisch et al. (2011), who found that listeners tended to be more strongly influenced by proximal rate than distal rate when the distal and proximal contexts were contradistinct.

This is not to say that the distal information manipulated in Experiment 1 *cannot* matter. The trend observed in Experiment 1 was largely in the expected direction, with slower distal speech rates leading to a lower proportion of “longer” responses to items in the critical region. It might be that the items used here were insufficiently ambiguous to allow the effects to attain significance. Dilley and Pitt (2010) were able to sample from a very wide variety of speakers to select the most acoustically ambiguous tokens of the items they used in their experiment, while we were constrained by the need to pair these items with segmentation ambiguity with analogous segment items. Perhaps if more strongly ambiguous individual items had been obtained, distal rate effects in Experiment 1 could have been present. Furthermore, Dilley and Pitt (2010) manipulated both preceding and following context, while this set of studies (and many segment studies) manipulated just preceding context. Some combination of these effects may have worked against finding distal rate effects in Experiment 1. Regardless, however, it seems safe to conclude that the effects of nearby context can alter people’s segmentation decisions, and that this may be playing a role in a number of studies exploring effects of “distal” rate information.

For segments, the word-initial findings are rather straightforward, while the word-final ones are less so. There was no evidence for distal effects on word-initial segment ambiguities across every definition of distal context. This matches the conclusions of a surfeit of studies looking at word-initial segments (Newman & Sawusch, 1996; Shinn et al., 1985; Summerfield, 1981). However, these other studies *did* generally find proximal effects, whereas ours did not; the reasons for this variation across studies is an area for future investigation, but may include the particular proximal manipulation that we chose to use in the present study, or acoustic properties of the particular stimuli that were used. In any case, distal information did not influence listeners’ perception of word-initial stop voicing in the materials selected here. For word-final segment ambiguities, meanwhile, there was evidence for distal rate effects for some, but not all, definitions of distal rate. Yet the pattern of results was rather counterintuitive: the conservative definition used in Experiment 1 and the difference definition of Experiment 3 led to significant distal effects, while the liberal definition of Experiment 2 (which was, in essence, a combination of the effects present in Experiment 1 and Experiment 3) did not lead to a change in word-final consonant perception. Although the reasons for the pattern of results by distal definition for word-final segment

ambiguities is baffling, the differences between word-initial and word-final ambiguities bear further inspection.

Word-final segment ambiguities are not common targets of speech rate experiments. Indeed, to our knowledge, this may be the first study to probe the effects of distal speech rate (under any definition) on word-final voicing ambiguities. And what was uncovered was that word-final ambiguities seem to be influenced by distal speech rate. Whether because the vocalic cue to word-final voicing has a similar duration to the segmentation ambiguities examined here, or the inconsistent palette of possible cues to word-final voicing making any one particular token more capable of being influenced by rate, distal rate was apparently perfectly capable of changing the perception of voicing at the end of a word.

This puts the word-initial segment contrasts in rather lonely company, as both segmentation and word-final voicing are capable of showing these distal rate effects. This suggests that typical questions about the provenance of distal rate effects on word-initial voicing might be better served by being flipped on their head. That is, rather than asking why, to the extent that it is, word-final voicing *can* be modulated by distal rate, a better question would be to ask why word-initial voicing *cannot* be modulated by it, given the possibility for rate adaptation effects in other ambiguities. One possible explanation is that rate ambiguities are simply not liable to the same strength of rate dependency in production (Nakai & Scobbie, 2016), which means that listeners generally do not attempt to use it in the course of perception. Such an explanation could benefit from studies of the production and perception of other phonetic ambiguities, such as the fricative–affricate contrasts that are also said to be relatively impervious to distal rate manipulations (Newman & Sawusch, 1996). Another possibility is that distal effects more strongly influence the perception of long segments (as with the vowels used to determine word-final voicing or the fricatives, nasals, and approximants that led to ambiguity to segmentation) than they influence the perception of the very short duration of VOT that signals word-initial voicing.

It therefore appears that at least two possibilities can hold water to explain previous patterns of results in the distal rate literature. First, the difference in definitions used did seem to make a difference in determining the strength of distal rate effects. For example, for segmentation ambiguities (both word-final and word-initial), distal rate effects largely seemed to be driven by speech information that was more than a syllable removed from a potential word boundary, but still relatively close to the point of maximal ambiguity. This neatly corresponds to the difference in definitions between studies primarily focusing on word segmentation and those primarily focusing on segment voicing. Furthermore, it appears that many of the differences that appeared in previous studies might also be the result of failing to include word-final segment ambiguities in studies of distal rate effects on voicing; although the present set of experiments involved replication of the lack of word-initial ambiguity effects, there was nevertheless

support for distal rate effects on word-final ambiguities. It does not seem to be necessary, then, to posit differences in representation or processing that would lead to the differences between segmentation and segment distal rate effects, but the present experiments also cannot rule such an explanation out entirely.

An interesting test of these claims would relate to languages that have length-based contrasts in their phonetic inventory. Arabic, for example, has vowels and consonants that have short and long realizations. These realizations are highly salient; several meaningful grammatical contrasts are conveyed only through changes in the duration of individual segments. However, the effects of distal prosodic cues on these contrasts are largely unknown. If distal prosodic effects are stronger on segmentation ambiguities than on segment ambiguities due to specialized processing mechanisms for segmentation, distal prosodic effects should be quite weak for all types of segments, including the perception of segment length. If, on the other hand, segment length contrasts are just as strongly affected by distal context as ones related to segmentation, this may argue for an explanation that does not require recourse to different processing streams. Preliminary results suggest that Arabic speakers' perception of consonant length is indeed dependent on distal rate.

A secondary objective of these studies in was the comparison between the online and offline participants. These model comparisons showed that there were no significant differences between participants recruited from UMD and participants recruited on Mechanical Turk for the segmentation items, but that word-final segment items were subject to differences in segment perception between UMD and Mechanical Turk. There was no interaction between participant group and distal rate, suggesting that the two factors operated on different facets of perception; the difference, then, was in the baseline likelihood to report hearing a word-final segment as voiced or voiceless.

Although we therefore lean toward endorsing the idea of using Mechanical Turk in phonetic perception experiments, caution should be used before throwing open the doors entirely. Some of the challenges of working with MTurk are not unique to experiments related to speech perception, or indeed to language. Although all of the local participants completed the task in somewhere between 20 and 30 minutes, the Mechanical Turk participants varied more, with times to completion ranging from 20 to 55 minutes. The MTurk participants who took almost an hour to participate in this study were almost certainly distracted by other tasks, or had poor Internet connections, possibilities not evident in person. Because the participants were not assigned randomly to the MTurk or UMD conditions, there may have been differences in the motivation and reward structure of the participant groups. However, some difficulties are more unique to the context of a speech perception experiment. For example, there was no way to know what headphones the participants wore, though some indicated that they failed to use headphones despite a specific request. These facts, both linguistic and

nonlinguistic, may have resulted in the different baselines uncovered for word-final segments; speech rate is a cue that is transmitted quite easily by almost all audio equipment, while, say, more subtle frequency distinctions might not be transmitted so easily. It is possible that these results would not generalize to all phonetic experiments.

In sum, we sought to examine the effects of far-away (distal) speech rate information on the perception of individual segments and word segmentation in speech. To compare the two possible percepts, a single methodology was established that would allow for simultaneous probing of both aspects of speech perception, both in person and using Amazon's Mechanical Turk service. Although these findings are complicated by a surprising failure to replicate proximal effects for these materials, it seems likely that at least two possibilities explain some of the differences found between previous studies of distal rate effects on segments and segmentation. First, the definition of "distal" (which differed between previous studies of segments and segmentation) influenced both the perception of segmentation ambiguities and of word-final segment ambiguities. Second, previous studies of distal rate effects on segments have only employed word-initial segment ambiguities. Although these studies' failure to find distal

rate effects was replicated, it also seems to be the case that word-final segment ambiguities in fact *are* influenced by distal speech rate. Further carefully designed studies will be necessary to tease apart these options and evaluate other alternatives.

**Acknowledgments** This work was supported by a National Science Foundation Graduate Research Fellowship award and a University of Maryland, College Park Graduate School Flagship Fellowship to CCH, as well as NSF IGERT Grant 0801465 and the Maryland Language Science Center. We thank Stephen DeVilbiss as well as Jessica Nwaogbe, Samantha Portuondo, and all the members of the UMD Language Development Lab for their help in running on-site participants for this project; Mikayla Abrams, Alison Arnold, Peter Enns, Maura O'Fallon, Ashley Thomas, and Eric Pelzl for recording experimental stimuli; and Lara Ehrenhofer, Stephanie Kramer, Natalia Lapinskaya, Chris Neufeld, Rachael Richardson, and Anton Rytting for their contributions to off-site experimental preparation and study running. Winter Storm participants more generally provided very useful feedback. Two anonymous reviewers gave key insights into the experiment here, including proposing the nucleus of Experiment 4 and suggesting that word-initial and word-final contrasts be probed separately from each other, both of which strongly improved the quality of this study. Scott Jackson provided important statistical expertise. Portions of this research were presented at the 21st International Congress on Acoustics in Montréal, Québec, travel to which was supported by the Department of Hearing and Speech Sciences at the University of Maryland, College Park.

## Appendix

Context	Segmentation	Segmental
The owner heard about the	four/fourth thefts.	fort/Ford thefts.
Bailey has much	bee/bean knowledge.	beet/bead knowledge.
The merchant sold	<sup>1,3,4</sup> Canadian oats/notes.	<sup>3</sup> Canadian coats/goats.
Becca makes use of the	<sup>3,4</sup> student's eye/sigh.	<sup>1,2</sup> student's pie/buy.
In the zoo, the buffalo	scare/scares zebra.	hurt/herd zebra.
The farmers bought some	<sup>1,2,4</sup> rye/rice somewhere.	<sup>1,2,3,4</sup> rope/robe somewhere.
I don't think Nate can	<sup>1,2,3,4</sup> see Mary/seem merry.	seat/seed Mary.
The channel cancelled the	new/noon newscast.	stunt/stunned newscast.
The tornado threatened the	bar/barn nearby.	dock/dog nearby.
The toddler smacked	their ears/rears.	<sup>1</sup> their tears/dears.
Sam ignored his	sock/socks splitting.	back/bag splitting.
It looks to me like this	<sup>3,4</sup> carpet's old/sold.	carpet's cold/gold.
Chris hates to assess the	tiger ages/rages.	tiger cages/gauges.
The business wanted the	<sup>2,3</sup> scarf ad/fad.	scarf patch/batch.
The travelers enjoyed the	<sup>3</sup> wharf air/fair.	<sup>1,2,3,4</sup> wharf path/bath.
To win he had to take	several aces/laces.	<sup>1,2,3,4</sup> several paces/bases.
The city must be saved from	its wine/swine.	its plight/blight.
Tigers really	<sup>1,2,3,4</sup> will earn/learn.	<sup>1,2,3,4</sup> will pounce/bounce.
She thought her foe was a	little affable/laughable.	<sup>1,3,4</sup> little tense/dense.
The housewife fought off	<sup>1,2,4</sup> Maine/Maine's zombies.	mop/mob zombies.
Peggy removed the	goo/goose swiftly.	threat/thread swiftly.
Georgia is known for	<sup>1,3,4</sup> its eels/seals.	its peaches/beaches.
Look for the woman whose hair	<sup>1,2,4</sup> looks light/slight.	<sup>1,2,4</sup> looks curly/girly.
Debra hates thinking about	her age/rage.	her tomb/doom.

We must rid the streets	of owls/vowels.	<sup>1,2,3,4</sup> of crime/grime.
The artist wouldn't want	<sup>1,2,3,4</sup> another evolution/revolution.	another calorie/gallery.
Her initials were	<sup>1</sup> S E/C.	S T/D.
The magician held out the	<sup>1</sup> die/dice skywards.	<sup>2</sup> cup/cub skywards.
Laura examined the	tea/teeth thoroughly.	lock/log thoroughly.
David asked if I'd seen his	play/place sometime.	<sup>1,2,3,4</sup> buck/bug sometime.
Doug scrutinized his	door/dorm more.	ballot/ballad more.
Kelsey needed some pig	<sup>1,2,3</sup> meat/meats Sunday.	feet/feed Sunday.
He didn't want to see his	brother ache/rake.	brother Ted/dead.
Ben wrote many	core/coarse sentences.	misspelt/misspelled sentences.
I wonder how they	<sup>4</sup> confer/confirm messages.	relate/relayed messages.
You can find it by the	<sup>1</sup> lime aisle/mile.	lime plume/bloom.
When doesn't Sean	<sup>1</sup> mow/moan nervously?	<sup>4</sup> write/ride nervously?
The actors sometimes	<sup>1,2,3,4</sup> draw/drawl lines.	<sup>4</sup> blurt/blurred lines.
He regretted his injured	<sup>1,2,4</sup> lie/life forever.	<sup>1,2,3</sup> sight/side forever.
It was a common	<sup>4</sup> Las Vegas inn/sin.	<sup>1,2,4</sup> Las Vegas pin/bin.
She was shocked by the state of	Pat's car/scar.	Pat's card/guard.
The lion was ashamed of	its kill/skill.	its plunder/blunder.
It's not a typical	Minneapolis ale/sale.	<sup>1,2,3,4</sup> Minneapolis pail/bail.
The museum had	numerous nails/snails.	<sup>1,4</sup> numerous trains/drains.
The spy eyed the	<sup>1,3</sup> President's peach/speech.	<sup>1</sup> President's palm/balm.
Nothing's more valuable than	<sup>2,4</sup> Vermont's oil/soil.	Vermont's tomes/domes.
I wouldn't call	<sup>4</sup> Olaf risky/frisky.	<sup>2</sup> Olaf classy/glassy.
Mark showed off his	<sup>1,3,4</sup> substantial ash/lash.	<sup>1,2,4</sup> substantial cash/gash.
I thought you would know how they	pray/praise zealously.	spelt/spelled zealously.
Please don't touch that	elephant/elephant's skin.	sunburnt/sunburned skin.
Put them there by the	<sup>2</sup> seven/seventh things.	<sup>1,2,3,4</sup> white/wide things.
Liz showed off her	<sup>4</sup> bow/bowl loudly.	cap/cab loudly.
The soldiers defended the	sergeant/sergeant's supplies.	eight/aid supplies.
At the state fair they	say/save vegetables.	grate/grade vegetables.
Put the peas in the	<sup>1,2,3,4</sup> soup/soups soon.	<sup>2,3</sup> pot/pod soon.
I really hate that	<sup>1,2,3,4</sup> impossible ump/lump.	impossible pump/bump.
The cop saw the	gun aim/name.	<sup>1,2,3</sup> gun came/game.
The teacher sketched a	firm arc/mark.	<sup>1,2,3,4</sup> firm pond/bond.
Sally had an	<sup>2,3</sup> endless nap/snap.	endless creed/greed.
Devin worries about	<sup>2,3,4</sup> cat/cats' snacks.	bat/bad snacks.
Christine took a	<sup>1,3,4</sup> serious spill/pill.	<sup>1,3,4</sup> serious punch/bunch.

Items with superscript numbers before them were excluded from analysis in that experiment because of entirely flat response functions in participants' responses.

## References

- Adank, P., & Janse, E. (2009). Perceptual learning of time-compressed and natural fast speech. *Journal of the Acoustical Society of America*, *126*(5), 2649–2659. doi:10.1121/1.3216914
- Baayen, R. H., Bates, D. M., Kliegl, R., & Vasishth, S. (2015). *RePsychLing: Data sets from psychology and linguistics experiments*. Retrieved from <https://github.com/dmbates/RePsychLing>
- 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
- Barr, D. J., Levy, R., Scheepers, C., & Tily, H. J. (2013). Random effects structure for confirmatory hypothesis testing: Keep it maximal. *Journal of Memory and Language*, *68*(3), 255–278. doi:10.1016/j.jml.2012.11.001
- Bates, D. M., Kliegl, R., Vasishth, S., & Baayen, H. (2015). Parsimonious mixed models. *ArXiv*. Retrieved from <https://arxiv.org/abs/1506.04967>
- Bates, D. M., Maechler, M., Bolker, B., & Walker, S. (2014). lme4: Linear mixed-effects models using Eigen and S4. Retrieved from <http://cran.r-project.org/package=lme4>

- Bell, A., Brenier, J. M., Gregory, M., Girand, C., & Jurafsky, D. (2009). Predictability effects on durations of content and function words in conversational English. *Journal of Memory and Language*, *60*(1), 92–111. doi:10.1016/j.jml.2008.06.003
- Bent, T., Bradlow, A. R., & Smith, B. L. (2008). Production and perception of temporal patterns in native and nonnative speech. *Phonetica*, *65*(3), 131–147. doi:10.1016/j.biotechadv.2011.08.021
- Boemio, A., Fromm, S., Braun, A., & Poeppel, D. (2005). Hierarchical and asymmetric temporal sensitivity in human auditory cortices. *Nature Neuroscience*, *8*(3), 389–395. doi:10.1038/nm1409
- Boersma, P., & Weenink, D. (2009). Praat: Doing phonetics by computer. *Glott International*. Retrieved from [www.praat.org](http://www.praat.org)
- Boucher, V. J. (2002). Timing relations in speech and the identification of voice-onset times: A stable perceptual boundary for voicing categories across speaking rates. *Perception & Psychophysics*, *64*(1), 121–130. doi:10.3758/BF03194561
- Breen, M., Dilley, L. C., McAuley, J. D., & Sanders, L. D. (2014). Auditory evoked potentials reveal early perceptual effects of distal prosody on speech segmentation. *Language, Cognition and Neuroscience*, *29*(9), 1132–1146. doi:10.1080/23273798.2014.894642
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in on-line sentence processing. *Psychonomic Bulletin & Review*, *18*(6), 1189–1196. doi:10.3758/s13423-011-0167-9
- Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2015). Metrical expectations from preceding prosody influence perception of lexical stress. *Journal of Experimental Psychology: Human Perception and Performance*, *41*(2), 306–323. doi:10.1037/a0038689
- Buhrmester, M., Kwang, T., & Gosling, S. D. (2011). Amazon's Mechanical Turk: A new source of inexpensive, yet high-quality, data? *Perspectives on Psychological Science*, *6*(1), 3–5. doi:10.1177/1745691610393980
- Clopper, C. G., & Smiljanic, R. (2015). Regional variation in temporal organization in American English. *Journal of Phonetics*, *49*, 1–15. doi:10.1016/j.wocn.2014.10.002
- Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America*, *83*(4), 1553–1573. doi:10.1121/1.395911
- Dilley, L. C., Mattys, S. L., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*(3), 274–294. doi:10.1016/j.jml.2010.06.003
- Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*(3), 294–311. doi:10.1016/j.jml.2008.06.006
- Dilley, L. C., Morrill, T. H., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalization. *Frontiers in Psychology*, *4*, 1002. doi:10.3389/fpsyg.2013.01002
- 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
- Dilley, L. C., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics*, *24*(4), 423–444. doi:10.1006/jpho.1996.0023
- Fujimura, O., Macchi, M. J., & Streeter, L. A. (1978). Perception of stop consonants with conflicting transitional cues: A cross-linguistic study. *Language and Speech*, *21*(4), 337–346.
- Goodman, J. K., Cryder, C. E., & Cheema, A. (2013). Data collection in a flat world: The strengths and weaknesses of Mechanical Turk sample. *Journal of Behavioral Decision Making*, *26*(3), 213–224. doi:10.1002/bdm.1753
- Heffner, C. C., Dilley, L. C., McAuley, J. D., & Pitt, M. A. (2013). When cues combine: How distal and proximal acoustic cues are integrated in word segmentation. *Language and Cognitive Processes*, *28*(9), 1275–1302. doi:10.1080/01690965.2012.672229
- Hickok, G., & Poeppel, D. (2007). The cortical organization of speech processing. *Nature Reviews Neuroscience*, *8*(5), 393–402. doi:10.1038/nm2113
- Hillenbrand, J., & Ingrisano, D. (1984). Perception of the voiced–voiceless contrast in syllable-final stops. *Journal of the Acoustical Society of America*, *76*(1), 18–26. doi:10.1121/1.391094
- Jacewicz, E., Fox, R. A., O'Neill, C., & Salmans, J. (2009). Articulation rate across dialect, age, and gender. *Language Variation and Change*, *21*(2), 233–256. doi:10.1017/S0954394509990093
- Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance*, *15*(4), 736–748. doi:10.1037/0096-1523.15.4.736
- Kleinschmidt, D., & Jaeger, T. F. (2012, August 1–4). A continuum of phonetic adaptation: Evaluating an incremental belief-updating model of recalibration and selective adaptation. *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (605–615). Sapporo, Japan: Cognitive Science Society.
- Kurumada, C., Brown, M., & Tanenhaus, M. K. (2012). Pragmatic interpretation of contrastive prosody: It looks like speech adaptation. In N. Miyake, D. Peebles, & R. P. Cooper (Eds.), *Proceedings of the 34th Annual Conference of the Cognitive Science Society* (pp. 647–652). Sapporo, Japan: Cognitive Science Society.
- Lai, W., & Dilley, L. C. (2016). Cross-linguistic generalization of the distal rate effect: Speech rate in context affects whether listeners hear a function word in Chinese Mandarin. In *Proceedings of Speech Prosody* (pp. 1124–1128). Boston, MA. doi:10.13140/RG.2.1.1749.4165
- Lisker, L., & Abramson, A. S. (1967). Some effects of context on voice onset time in English stops. *Language and Speech*, *10*(1), 1–28. doi:10.1177/002383096701000101
- Marge, M., Banerjee, S., & Rudnicky, A. I. (2010). Using the Amazon Mechanical Turk for transcription of spoken language. In J. Hansen (Ed.), *Proceedings of the 2010 I.E. Conference on Acoustics, Speech and Signal Processing* (pp. 5270–5273). IEEE. doi:10.1109/ICASSP.2010.5494979
- Mattys, S. L. (1997). The use of time during lexical processing and segmentation: A review. *Psychonomic Bulletin & Review*, *4*(3), 310–329. doi:10.3758/BF03210789
- Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General*, *134*(4), 477–500. doi:10.1037/0096-3445.134.4.477
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*(1), 1–86. doi:10.1016/0010-0285(86)90015-0
- Miller, J. L. (1994). On the internal structure of phonetic categories: A progress report. *Cognition*, *50*(1/3), 271–285. doi:10.1016/0010-0277(94)90031-0
- Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance*, *14*(3), 369–378. doi:10.1037//0096-1523.14.3.369
- Miller, J. L., Green, K. P., & Reeves, A. (1986). Speaking rate and segments: A look at the relation between speech production and speech perception for the voicing contrast. *Phonetica*, *43*(1/3), 106–115. doi:10.1159/000261764
- Miller, J. L., & Grosjean, F. (1981). How the components of speaking rate influence perception of phonetic segments. *Journal of Experimental Psychology: Human Perception and Performance*, *7*(1), 208–215. doi:10.1037/0096-1523.7.1.208
- Miller, J. L., & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics*, *25*(6), 457–465. doi:10.3758/BF03213823

- Miller, J. L., & Volaitis, L. E. (1989). Effect of speaking rate on the perceptual structure of a phonetic category. *Perception & Psychophysics*, 46(6), 505–512. doi:10.3758/BF03208147
- Nakai, S., & Scobbie, J. M. (2016). The VOT category boundary in word-initial stops: Counter-evidence against rate normalization in English spontaneous speech. *Laboratory Phonology*.
- 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
- Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, 52(3), 189–234. doi:10.1016/0010-0277(94)90043-4
- Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review*, 115(2), 357–395. doi:10.1037/0033-295X.115.2.357
- Peirce, J. (2007). PsychoPy—Psychophysics software in Python. *Journal of Neuroscience Methods*, 162(1/2), 8–13. doi:10.1016/j.jneumeth.2006.11.017
- Poeppl, D., Idsardi, W. J., & van Wassenhove, V. (2008). Speech perception at the interface of neurobiology and linguistics. *Philosophical Transactions of the Royal Society of London B: Biological Sciences*, 363(1493), 1071–1086. doi:10.1098/rstb.2007.2160
- Quené, H., & van den Bergh, H. (2008). Examples of mixed-effects modeling with crossed random effects and with binomial data. *Journal of Memory and Language*, 59(4), 413–425. doi:10.1016/j.jml.2008.02.002
- Raphael, L. J. (1972). Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English. *Journal of the Acoustical Society of America*, 51(4, Pt. 2), 1296–1303. doi:10.1121/1.1912974
- 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
- Repp, B. H. (1978). Perceptual integration and differentiation of spectral cues for intervocalic stop consonants. *Perception & Psychophysics*, 24(5), 471–485.
- Repp, B. H. (1982). Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception. *Psychological Bulletin*, 92(1), 81–110. doi:10.1037/0033-2909.92.1.81
- 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
- Schmidt, A. M., & Flege, J. E. (1995). Effects of speaking rate changes on native and nonnative speech production. *Phonetica*, 52(1), 41–54.
- Schouten, M. E. H., & Pols, L. C. W. (1983). Perception of plosive consonants: The relative contribution of bursts and vocalic transitions. In M. von den Broecke, V. J. van Heuven, & W. Zonneveld (Eds.), *Sound structures: Studies for Antoine Cohen* (pp. 227–243). Dordrecht, Germany: Foris Publications.
- Schwab, S., & Avanzi, M. (2015). Regional variation and articulation rate in French. *Journal of Phonetics*, 48, 96–105. doi:10.1016/j.wocn.2014.10.009
- Seidl, A. (2007). Infants' use and weighting of prosodic cues in clause segmentation. *Journal of Memory and Language*, 57(1), 24–48. doi:10.1016/j.jml.2006.10.004
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25(2), 193–247. doi:10.1007/BF01708572
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, 68(1), 1–16. doi:10.3758/BF03193651
- Shinn, P. C., Blumstein, S. E., & Jongman, A. (1985). Limitations of context conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics*, 38(5), 397–407. doi:10.3758/BF03207170
- Slote, J., & Strand, J. F. (2016). Conducting spoken word recognition research online: Validation and a new timing method. *Behavior Research Methods*, 48(2), 553–566. doi:10.3758/s13428-015-0599-7
- Sprouse, J. (2011). A validation of Amazon Mechanical Turk for the collection of acceptability judgments in linguistic theory. *Behavior Research Methods*, 43(1), 155–167. doi:10.3758/s13428-010-0039-7
- 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. (2014, September). The time-course of speaking rate compensation: Effects of sentential rate and vowel length on voicing judgments. *Language, Cognition and Neuroscience*, 1–15. doi:10.1080/23273798.2014.946427
- Verhoeven, J., De Pauw, G., & Kloots, H. (2004). Speech rate in a pluricentric language: A comparison between Dutch in Belgium and the Netherlands. *Language and Speech*, 47(3), 297–308. doi:10.1177/00238309040470030401
- Volaitis, L. E., & Miller, J. L. (1992). Phonetic prototypes: Influence of place of articulation and speaking rate on the internal structure of voicing categories. *Journal of the Acoustical Society of America*, 92(2, Pt. 1), 723–735. doi:10.1121/1.403997
- Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America*, 95(5, Pt. 1), 2694–2701. doi:10.1121/1.409838