Atten Percept Psychophys (2016) 78:334–345
DOI 10.3758/s13414-015-0981-7

Rate dependent speech processing can be speech specific: Evidence from the perceptual disappearance of words under changes in context speech rate

Mark A. Pitt[1] · **Christine Szostak**[1] · **Laura C. Dilley**[2]

Published online: 22 September 2015

Abstract The perception of reduced syllables, including function words, produced in casual speech can be made to disappear by slowing the rate at which surrounding words are spoken (Dilley & Pitt, Psychological Science, *21*(11), 1664–1670. doi: 10.1177/0956797610384743, 2010). The current study explored the domain generality of this speech-rate effect, asking whether it is induced by temporal information found only in speech. Stimuli were short word sequences (e.g., *minor or child*) appended to precursors that were clear speech, degraded speech (low-pass filtered or sinewave), or tone sequences, presented at a spoken rate and a slowed rate. Across three experiments, only precursors heard as intelligible speech generated a speech-rate effect (fewer reports of function words with a slowed context), suggesting that rate-dependent speech processing can be domain specific.

Keywords Speech rate · Spoken word recognition · Domain generality · Phonetic perception

The perception of speech requires sensitivity to precise timing over multiple time scales. To identify and discriminate phonemes, listeners must be sensitive to differences in voice onset time (VOT) (Miller 1981; Port 1979), segment duration, and relative cue timing (e.g., trading relations; Best et al. 1981). To perceive lexical stress and syllabify words, listeners must be sensitive to durational differences across syllables (Reinisch et al. 2011a, b; Turk and Sawusch 1997; Turk and Shattuck-Hufnagel 2000). Although fluid communication requires the ability to perceive speech at different rates, little work has directly explored how speech rate contributes to the perception of spoken words.

Dilley and Pitt (2010) argued that speech rate can be a valuable cue in spoken word recognition because it can partially compensate for the absence of other cues when the talker's speech is highly reduced, such as when speaking in a casual style. Function words (e.g., *of*, *or*, *in*) are particularly vulnerable to distortion because they are heavily coarticulated with surrounding words and can be very short in duration (50 ms). In particular, when the phonemes of a function word match those in the rhyme of the preceding word (e.g., *minor or*), the two words can blend together when heavily coarticulated, creating what can be considered an elongated production of the first word (e.g., *minorrr*). When looked at spectrographically, the words are spectrally indistinct, with no changes in frequency or amplitude that would normally signal a word boundary. For this reason, Dilley and Pitt (2010) argued that timing information from the surrounding context is a crucial cue that listeners use to perceive short function words. That is, context speech rate assists in determining whether the talker said *minor* or *minor or*.

In an experimental set-up designed to elicit casual-style speech, Dilley and Pitt (2010) had talkers produce sentences containing two-word sequences that are prone to such blending (e.g., *Anyone must be a minor or child to enter*). They then took these productions and varied the speech rate of a small (critical) region of the sentence containing the function word (e.g., *-nor or ch-*) and the remainder of the sentence (preceding and following words). In the two conditions of primary interest for the current study, the critical (proximal) region was held constant and the rate of the distal context (precursor) was varied, being presented at the rate spoken by the talker

✉ Mark A. Pitt
pitt.2@osu.edu

[1] Department of Psychology, The Ohio State University, 1835 Neil Avenue, Columbus, OH 43220, USA

[2] Department of Communicative Sciences and Disorders, Michigan State University, 1026 Red Cedar Road, East Lansing 48824-1220, MI, USA

or time expanded by a factor of 1.9. Listeners heard the sentences over headphones and were instructed to type exactly what was heard using the computer keyboard.

In the spoken-rate condition, listeners reported the function word 79 % of the time. Function word reports dropped to 33 % when the context was slowed. A similarly low report of function words (35 %) was found when the speaking rate of the critical region was sped up (using speech compression), while presenting the precursor at the spoken rate. Finally, their 2 showed that changes in speech rate can induce listeners to report function words that the talker never said. For example, when presented with a casual production of the sentence *Anyone must be a minor child*, listeners reported a function word (e.g., *or, and*) between *minor* and *child* 24 % of the time when the speech rate of the precursor, but not the critical region (e.g., *-nor ch-*), was sped up by a factor of 0.6. The fast rate of the distal context caused listeners to infer that the talker spoke an extra syllable in the critical region.

Subsequent studies have probed various questions about this lexical rate effect (LRE). Heffner et al. (2013) examined the interaction among the distal and proximal cues by varying the strength of the acoustic cues (intensity, F0, duration) specifying the function word and the strength of rate cue in the precursor. The results showed that word duration interacted with speech rate in influencing function word reports, whereas intensity and F0 tended to combine with speech rate more additively. Notably, the results showed that the more immediate proximal acoustic cues do not simply override the more distal cue of speech rate. Both contribute to how the critical region is perceived. Importantly, this paper additionally showed that the strength of the LRE varies continuously (and linearly) as a function of distal speech rate.

Other studies have examined the generalizability of the LRE. Dilley et al. (2013) demonstrated the LRE in a different morphosyntactic environment by replicating Dilley and Pitt (2010) in Russian. Significantly, their data suggested that the LRE is not specific to function words but instead applies more generally to reduced syllables. In addition, they reported preliminary evidence that language experience is positively related to the strength of the LRE. Morrill et al. (2014) showed that the LRE can be obtained by varying not just the rate of speech but also the rhythm of the distal context (binary vs. ternary patterns). Listeners were more likely to report a function word when the rhythmic organization of the critical region matched that of the context. As in Heffner et al. (2013), they also found that effects of rate and rhythm were additive, with function word reports being highest when rate and rhythm reinforced the same interpretation of the critical region.

Most recently, Baese-Berk et al. (2014) showed that the timeframe over which the LRE occurs is not limited to the immediately preceding sentence frame (second or two of speech) but rather can build over an hour. Listeners heard 189 sentences that formed a distribution of speech rates. Three groups of listeners received distributions that varied in mean speech rate (e.g., 1.2, 1.4, 1.6 times the spoken rate). Across the course of the experiment, listeners' reports of function words were increasingly influenced by the distribution of sentences they heard, not just by the rate of the sentence on a given trial. By the end of the experiment, although all three groups responded to a subset of sentences at the 1.4 speaking rate, function word reports were 22 % more frequent by listeners whose mean speech rate was the slowest (1.2) than the fastest (1.6).

This group of studies demonstrates that the LRE is a robust perceptual phenomenon, replicating across studies and tasks (Heffner et al. 2013), and generalizing to another language. The studies also show that the LRE provides a window into the properties of a timing mechanism involved in the perception of speech, revealing that the pace of speech, established over tens of minutes, and the rhythm of speech (prosody) can alter perception.

Our current explanation of the LRE is that it reflects violations of temporal expectations established by the rate of the distal speech. We hypothesize that the brain entrains to speech rate upon hearing it. Entrainment is a synchronization process by which the perceptual system is engaged to encode speech at the talker-produced rate, one by-product of which is the establishment of expectations in the processing system (i.e., making predictions) about the rate at which future speech is likely to be produced and therefore should be encoded. Entrainment simultaneously serves to synchronize perceptual encoding with the current rate of speech as well as maintain this coupling by forecasting the likely rate of upcoming speech (e.g., next few syllables).

The LRE occurs because the perceptual system is fooled into processing the critical region at the wrong rate. Take the situation in which the surrounding context is time-expanded and the critical region is at the spoken rate. The perceptual system entrains to the slow rate of the context and forecasts the critical region to be at this same rate. However, because the rate of the critical region is almost double that of the context, it is reinterpreted at the slower, distal rate, causing what is essentially the loss of a syllable (*minor or child* → *minor child*). This reinterpretation is possible because the spectral detail in the critical region is a viable reinterpretation at the context speech rate. If it were not, perception would be disrupted.

Listeners can entrain to other types of auditory events beside speech, such as music and environmental sounds (Large and Jones 1999). It is therefore of interest to know whether the entrainment mechanism responsible for the LRE is at all dependent on the context being speech or whether other types of sounds (e.g., tones) can be equally effective at inducing the LRE. By exploring this issue, we can learn whether the processes engaged in tracking rate are at all specific to the stimulus being tracked. An answer to this question can assist in

elucidating the conditions in which the LRE emerges and thus assist in explaining this perceptual phenomenon.

The purpose of this study was to evaluate the domain generality of the LRE. In three experiments, precursors of different types (clear speech, tones, filtered speech, sine-wave speech) were compared on their ability to produce an LRE. If the timing information in the precursor that conveys rate is not unique to speech, similar results should be found across all types of precursors. Alternatively, if the information necessary for conveying rate is domain specific, then selectivity should be found, with only speech precursors yielding an LRE. The various types of speech precursors were included to determine more precisely the conditions necessary to obtain an LRE.

Experiment 1

Three precursors were compared on their ability to alter the frequency of function word reports when the precursor was presented at two rates of speech—a spoken rate and a slowed rate. A clear speech precursor condition was included to demonstrate that an LRE could be obtained with the stimuli and to serve as a reference for comparison of rate effects across the other conditions. A tone precursor condition was included to test whether the rate effect is domain general. Tone precursors were chosen because their rate of presentation has been shown to influence labeling of steps along a phonetic continuum. For example, Wade and Holt (2005) reported that a rapid, isochronous sequence of sine tones induced more /ba/ responses along a /ba/–/wa/ continuum, whereas a slow sequence led to more /wa/ responses.

In the third condition, the speech precursor was low-pass filtered. Its purpose was two-fold. One was to provide a further test of the idea that, like tones, rate information is contained in the low-frequency, time-varying properties of speech. The second purpose was to assist in interpreting a possible null result in the tone condition. If a rate effect is found with filtered speech but not tones, the results would suggest that tone sequences lack the necessary timing information to convey speech rate. However, if even the filtered precursor fails to yield a rate effect, the results across conditions would suggest that only clear speech contains the timing information necessary to alter listeners' reports of function words.

Method

Participants

Seventy-two speakers of American English with reported normal hearing participated in exchange for course credit (24 in each precursor condition).

Design

A factorial design was used in which precursor type (clear speech, tone, filtered speech) was manipulated between subjects and speech rate (spoken rate, slowed rate) was manipulated within subjects.

Stimuli

Stimuli were 48 sentence fragments from those collected in the production experiment of Dilley and Pitt (2010). They were spoken by male and female talkers, and the digital recordings were stored at a sampling rate of 22.05 kHz with 16-bit quantization. Sentences consisted of two parts: a critical region, which contained the function word, and a precursor region, which corresponded to all words preceding the critical region. The critical region included the function word, the word preceding it, and the onset of the word following the function word (e.g., *Anyone must be a minor or child*, where the underlined portion represents the critical region). The word whose onset was part of the critical region (e.g., *child* in this example) was the last word in the sentence. When the word preceding the function word was disyllabic (e.g., *minor*), the critical region included one more syllable (e.g., *min* in *minor*) than the critical region in Dilley and Pitt (2010). This change was necessary to avoid only the second syllable being clear speech when the precursor was, for example, replaced by tones, which would have resulted in the first syllable being a tone.

The sentences in their naturally produced form served as the stimuli in the spoken-rate condition of the clear precursor condition. To create the stimuli for the corresponding slowed-rate condition, the precursor was time-expanded by a factor of 1.9, following Dilley and Pitt (2010), and the critical region was left at its spoken rate. Although the speech sounded slow, the expanded sentences were highly intelligible.

For the tone precursor condition, tone construction followed the general methods of two comparable studies that examined rate effects in phonetic labeling (Gordon 1988; Summerfield 1981), in which tone sequences were customized to match temporal and rhythmic properties the speech precursors. Sequences of seven-harmonic complex tones were created with a fundamental frequency of 110 Hz. Tone duration was 100 ms with 10 ms on/off ramps. The number of tones in each tone precursor matched the number of syllables in a corresponding clear speech precursor, and the temporal onset of each tone was adjusted to match that of the corresponding syllable. Thus, each tone precursor matched the duration and the rhythm of its corresponding speech precursor. This method of precursor construction was used for the spoken-rate condition. For the slowed-rate condition, the tones and the intertone intervals were lengthened by a factor of 1.9, holding pitch constant. Tone precursors were then matched to their respective critical regions in amplitude and prepended to the critical regions.

To create the stimuli for the filtered precursor condition, the precursors of the stimuli in the clear condition (spoken-rate and slowed-rate items) were low-pass filtered at 3.5 times the mean F0 of the talker who spoke the sentence. Across talkers, the frequency cut-off ranged from 391 Hz to 996 Hz. Use of a filter cut-off frequency that is linked to F0 helped standardize the acoustic information (e.g., harmonics) in the stimuli across talkers, who varied greatly in F0 (talkers were male and female undergraduates), but in no way left the speech intelligible. Across talkers, the filtering method ensured that F1, but not higher formants, was in the passband for sonorant segments (Guenther et al. 1999; Hillenbrand et al. 1995), while all high-frequency energy necessary for distinguishing consonants was in the stopband. The filtered precursors sounded like muffled speech, but were completely unintelligible. The filtered precursors were then matched to their critical regions in amplitude before being prepended to them. The low-pass filtered precursors at the original rate of speech served as the stimuli in the spoken-rate condition. For the slowed-rate condition, the filtered stimuli were lengthened by a factor of 1.9, holding pitch constant.

Procedure

Within each precursor condition, stimuli were counterbalanced across two lists, each containing 24 spoken-rate and 24 slowed items, in a manner that ensured the two versions of the same sentence did not occur in the same list. Stimuli within a list were presented in a pseudo-random order.

Participants were tested individually in a sound-dampened room. They were seated in front of a computer and heard all stimuli over headphones at a comfortable listening level. A trial began with presentation of a stimulus. Upon its offset, participants had unlimited time to type the sentence using a computer keyboard. The text that was typed appeared on a computer monitor, and participants could revise responses. Participants pressed the "Enter" key to signal they were finished responding. After a1-s pause, the next trial began. Six practice trials preceded the test trials, which together lasted approximately 15 minutes.

Results and Discussion

Responses were scored by determining on each trial whether the participant typed the function word in the critical region. If a function word was typed between two adjacent content words (e.g., *minor or child*), we inferred that listeners heard a function word and scored the trial as 1. If a function word, or any word, was not typed between the two adjacent words (e.g., *minor child*), we inferred listeners did not perceive a function word, and scored the trial as 0 (additional details on scoring can be found in Dilley and Pitt 2010).

The proportion of function words that listeners reported in the spoken-rate and slowed condition is shown in Fig. 1a as a function of each type of precursor. When the precursor was clear speech, a sizeable effect of speaking rate was obtained, with function word reports being .77 in the spoken-rate condition and dropping to .47 in the slowed-rate condition. In contrast, in the tone precursor condition, no effect of speech rate was obtained, with function word reports being nearly identical at the two speaking rates. In the filtered precursor condition, a similar null effect of speaking rate was obtained: Reports of the function word in the spoken-rate condition differed from those in the slowed condition by only -.02.

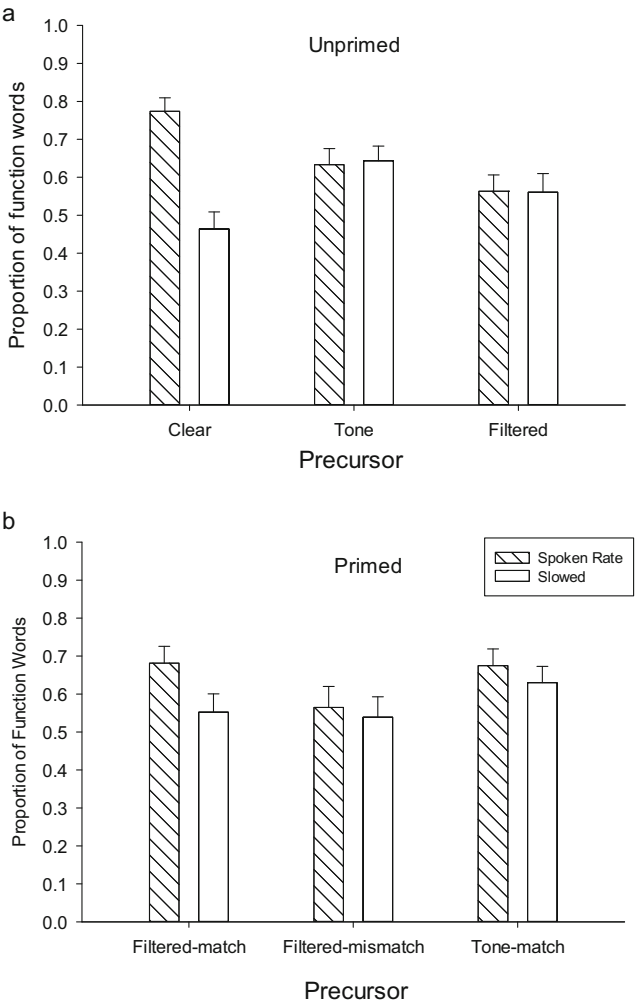

Fig. 1 Function word reports as a function of precursor and speech rate conditions. The upper graph (**a**) contains the data from Experiment 1, and the lower graph (**b**) the data from Experiment 2. Error bars represent 95 % confidence intervals

Statistical analyses confirmed what is evident in the graph. A mixed-models logistic regression analysis was performed on the data using the lme4 package (Bates et al. 2012) in the R statistical software (Version 2.13; the R Foundation for Statistical Computing), treating speech rate and precursor type as fixed factors. Preliminary analyses showed that the most appropriate random factors included participant intercepts and item intercepts and slopes. Likelihood ratio tests showed that

the saturated model was the best fitting model, with the interaction of speech rate and precursor type being reliable ($b = 0.95$, $p < .001$). Planned comparisons for each precursor type showed that the rate manipulation was reliable only for the clear speech precursor ($b = 0.30$, $p < .001$; tone: $b = -0.01$, $p < .95$; filtered: $b = 0.01$, $p < .95$). The LRE in the clear speech condition was reliably larger than that in the tone and the filtered precursor conditions (tone: $b = 1.09$, $p < .001$; filtered: $b = 1.90$, $p < .001$).

The results of Experiment 1 provide preliminary evidence that the LRE is specific to speech. If it were not, and the LRE instead reflected a domain-general entrainment mechanism, tone precursors should have altered function word reports, but there was no hint that this occurred. These data suggest that clear speech conveys timing information that is missing from multiharmonic tones, and which is critical for observing the LRE. Somewhat unexpectedly, filtered precursors also failed to alter listeners' reports of function words and suggests that only clear speech contains the timing information necessary to alter listeners' reports of function words.

The null effect of rate with tone precursors can be difficult to interpret because it could be due to peculiarities of the tones themselves and have nothing to do with conveying rate. This same concern influenced the design of the experiments by Wade and Holt (2005) when examining rate-based effects on phonetic labeling. Perhaps tones created in a different manner would yield an LRE? We addressed this question by rerunning the tone-precursor condition twice more. In the first follow-up test, we attempted to make the tones much more speech-like than those in Experiment 1. A tone sequence was stylized for every precursor such that tone durations matched the durations of intervals between vowel onsets of successive syllables; moreover, each tone had a static or gliding F0 that was based on the expected perceived pitch for the syllable (d'Alessandro and Mertens 1995) as implemented in Prosogram software (Mertens, 2004). For the second follow-up test, the tone sequences were partially patterned after those of Wade and Holt (2005). Because they obtained a rate effect with tones, it seemed reasonable to mimic aspects of their set-up. Isochronous sequences of short and long tones were created according to the durational specifications in their first experiment. However, instead of having these tones be sinewaves that vary randomly in frequency from tone to tone, as they did, the frequency composition of the tones was held constant, being the single multiharmonic tone used in Experiment 1. The data from both follow-up experiments resemble the tone data in Fig. 1a[1] Precursor rate did not differentially affect function word reports. The consistent null effect of speech rate across a total of three tone precursor conditions strengthens the argument that tones lack the rate information necessary to alter function word reports and reinforce the conclusion that the LRE reflects an aspect of temporal processing that is induced only by speech.

The failure to obtain a speech-rate effect in the filtered precursor condition was surprising because low-pass filtered speech contains prosodic cues that have been shown to influence rate and speech perception in multiple ways. For example, low-pass filtered speech conveys language-specific rhythmic properties (Mehler et al. 1988). Moreover, low-frequency temporal cues associated with the amplitude envelope of speech enable listeners to understand spectrally degraded speech so long as sufficient spectral detail is retained, particularly when they are trained on the degraded speech (Baer and Moore 1993; Davis et al. 2005; Elliott and Theunissen 2009; Peelle and Wingfield 2005; Shannon et al. 1995). Conversely, when low-frequency temporal cues below about 16 kHz are removed, the intelligibility of speech is substantially degraded (Drullman et al. 1994a, b; Ghitza 2012).

It may be that the slow, time-varying cues in filtered speech (amplitude envelope, F0) are not sufficient to convey the timing information necessary to generate a difference in function word reports (i.e., the LRE). The data of Experiment 1 suggest that more precise temporal information appears to be needed, which must be contained in the energy at higher frequencies because the clear and filtered conditions differed only in this property. Of course, the presence of higher frequency energy also makes speech intelligible, so it may be that phonetic intelligibility itself is necessary to observe the LRE. By disambiguating the precursor so that listeners perceive it as the talker intended (i.e., mentally restoring the phonetic detail, but crucially at a particular rate), an LRE might emerge. We explored this idea in Experiment 2 as a further test of the domain generality of the LRE.

Experiment 2

To examine whether intelligibility is necessary to generate an LRE, we used a priming methodology in which a clear precursor served as the prime for the corresponding filtered precursor stimulus from Experiment 1. In the filtered-match condition, the prime was the clear (i.e., unfiltered) version of the corresponding filtered precursor. In the filtered-mismatch condition, the prime was a clear, unfiltered precursor from another phrase. If the hypothesis concerning the ambiguity of the filtered speech is correct, the prime should disambiguate the precursor in the filtered-match condition, yielding an LRE. No such effect of rate should be found in the filtered-mismatch condition because the prime would not disambiguate the words in the precursor.

The priming methodology also provided another opportunity to test domain generality. If priming is found in the filtered-match condition, one might wonder whether any prime-precursor condition pairing that matched (on some dimension) could yield an LRE. In particular, what if a speech prime were paired with a precursor that was a sequence of

[1] Graphs of these data can be found at lpl.psy.ohio-state.edu/publications.php

tones derived from the original clear speech, as in the tone condition of Experiment 1? An LRE in this tone-match condition would be evidence of domain generality. No effect of speech rate in this tone-match condition would reinforce the findings of domain specificity in Experiment 1.

Method

Participants

Participants were 60 new individuals from the same population as Experiment 1 (20 in each precursor condition).

Design

Type of precursor (filtered-match, filtered-mismatch, tone-match) was manipulated between subjects and speech rate (spoken-rate, slowed) was manipulated within subjects.

Stimuli

The stimulus pairings were created by using a given clear speech precursor from the clear condition of Experiment 1 as the prime and one of the stimuli from Experiment 1 as the precursor-critical region sequence. To create the filtered-match stimuli, a clear speech prime was paired with its corresponding low-pass filtered precursor-critical region sequence. In this condition, the prime and precursor were identical except that the latter was filtered. To create the filtered-mismatch stimuli, these same primes and precursor-critical sequences were re-paired so that they mismatched; that is, the prime was not an unfiltered version of the precursor. Every effort was made to ensure that the replacement prime matched the original prime in length, syllable and word count, stress pattern across words, and semantic plausibility/syntactic legality of continuation into the critical region. On this last dimension, the quality of the pairing ranged from a plausible substitute to semantically/syntactically illegal/implausible. The stimuli for the tone-match condition were identical to those of the filtered-match condition except that the precursor-critical sequences were those from the tone condition of Experiment 1. In all three conditions, the speech rate of the prime matched that of the precursor (i.e., spoken-rate and slowed primes were prepended to spoken-rate or slowed precursors, respectively).

Procedure

The testing procedure was identical to that of Experiment 1 with a few modifications. Each trial began with presentation of a prime, a 2-s pause, and then the corresponding precursor-critical sequence. Participants were instructed to type the words spoken in the precursor-critical region only (not the prime), and they were told that attending to the prime could help them understand the precursor.

Results and Discussion

The data were scored and analyzed following the procedure of Experiment 1. The proportion of function words reported in the six conditions is shown in Fig. 1b. Overall, the results reinforce those of Experiment 1 in arguing that the precursor must be intelligible in order to convey rate information sufficient to produce an LRE. There is an effect of speech rate in the filtered-match condition, although it is smaller than that found in Experiment 1. Function word reports were .13 higher when the precursor was presented at the spoken rate than at the slowed rate. This LRE dropped to .02 when the prime mismatched the filtered target. In the tone-match condition, function word reports in the slowed-rate condition were .04 less than those in the spoken-rate condition.

The results of a mixed-models analysis with precursor condition and speech rate as fixed factors yielded a marginally reliable interaction of the two variables (b = -.236, p < .07). Comparisons of the effect of rate within each precursor condition proved reliable only in the filtered-match condition (b = -0.127, p < .001; filtered-mismatch: b = -0.250, p < .85; tone-match: b = -0.04, p < .60). The LRE in the filtered-match condition was reliably larger than that in the filtered-mismatch condition (b = -0.703, p < .01) but only approached significance in the tone-matched condition (b = -0.253, p < .11).

The logic of the priming manipulation hinges on the prime disambiguating the precursor when the prime and precursor matched but not when they mismatched. If this did not occur, we cannot infer that the difference between the two conditions is due to the relationship between the prime and precursor. To evaluate the effectiveness of the matching manipulation in disambiguating the precursor in the filtered-match and filtered-mismatch conditions, we compared the accuracy with which listeners transcribed the precursor. The accuracy of exact word transcriptions was scored. Mean transcription accuracy was 71 % vs. 4 % in the filtered-match and filtered-mismatch conditions, respectively, confirming that priming achieved its intended goal of disambiguating the precursor as intelligible speech only in the filtered-match condition. In the tone-match condition, listeners hardly ever typed words to represent the precursor tones.

Additional evidence demonstrating the selectivity of priming comes from a comparison of the LRE across conditions. The LRE cannot be due to the prime directly influencing perception of the critical region, irrespective of the precursor. Otherwise a comparable LRE should have been found in the tone-match condition. In addition, the LRE observed in the filtered-match condition must be due to the prime disambiguating the precursor. Otherwise an LRE should have been found in the filtered condition of Experiment 1.

The results of Experiment 2 show that a low-pass filtered precursor can produce an LRE, but in order to do so the precursor must be disambiguated to convey the words from which it was derived. In the filtered-match condition, the prime disambiguated the phonetic content of the precursor, turning it into a string of somewhat intelligible words. A consequence of perceiving the precursor as (intelligible) speech is that the timing relations among syllables and the segments within them are clarified. This added temporal precision, when supported by knowledge of the linguistic structures evoked by an intelligible signal, seems to provide the necessary information to generate the LRE. This finding, along with the failure to observe a comparable priming result when tones were precursors, further demonstrates the selectivity of this perceptual phenomenon. Specific conditions must be met to alter function word reports, and a particularly salient condition appears to be that the precursor must be heard as intelligible speech.

Across Experiments 1 and 2, the size of the LRE fell from .30 in the clear precursor condition of Experiment 1 to .13 in the filtered-match condition of Experiment 2. The reason for the drop appears to be two-fold. One is that function word reports in the spoken-rate condition decreased from .78 to .68. This drop is due primarily to listeners in Experiment 2 occasionally reporting a precursor that was not syntactically felicitous with the occurrence of a function word in the critical region (e.g., for the stimulus item *The snow suits are our only*, where the underlined portion represents the critical region, responses such as *This says he is our only* would not remain syntactically felicitous for a function word to be inserted within the underlined region). When these items are removed from the analysis, function word reports increase by .06 to .74. The second reason for the drop in the size of the LRE in Experiment 2 is that in the slowed-rate condition, function word reports *rose* .08 from Experiment 1 to Experiment 2. This rise indicates a reduction in the effectiveness of the slowed context in inducing listeners to perceive one less syllable in the critical region, and would seem to be due to the precursor being low-pass filtered.

Experiment 3

The purpose of Experiment 3 was to strengthen the finding in Experiment 2 that intelligible speech is required to generate the LRE. A limitation of that experiment is that physical differences between the speech and tone precursors could have contributed to or caused the LRE, and not, as we have argued, differences in how the precursor were themselves processed (as intelligible speech or tones). To eliminate this alternative interpretation, we minimized stimulus differences in Experiment 3 by using sinewave speech (Remez et al. 1981), which can be heard as (intelligible) speech or (unintelligible) nonspeech, depending on instruction or with minimal changes to the stimulus. Two sinewave versions of the stimuli in Experiment 1 were created. Those in the original condition contained precursors that were typical sinewave analogs. Precursors in the flipped condition were identical except that the F2 sinewave was flipped (spectrally rotated) along the time axis, rendering the speech unintelligible. The LRE should be found only in the original condition if the phenomenon depends on the precursor being heard as intelligible speech.

The second goal of Experiment 3 was methodological. Heffner et al. (2013; see also Dilley et al. 2013) used a two-choice classification task to study the LRE, having listeners report whether the critical region matched one of two possibilities (e.g., *minor* vs. *minor or*). Sizeable effects of speech rate on labeling were found. We used the same set-up here to generalize the specificity results using another task.

Method

Participants

Thirty-two individuals from the same pool and meeting the same criteria as those in Experiment 1 served as participants.

Design

The type of sinewave precursor (original, flipped) was manipulated between subjects and speech rate (spoken rate, slowed rate) was manipulated within subjects.

Stimuli

Stimuli were four sentences from the clear-speech condition of Experiment 1. They were chosen because their sinewave replicas were some of the most intelligible, which was determined in a pilot experiment. Sinewave versions of the spoken-rate and slowed stimuli were created using the Praat script written by Chris Darwin and then hand altering sinewave components to increase intelligibility and naturalness. These items served as the stimuli in the original condition. Stimuli in the flipped condition were created by spectrally rotating the F2 component of the precursor in the original stimuli; rotation was around the mean F2 frequency, so that the sinewave component roughly stayed in place in the frequency domain. Final portions of the sinewave components of the flipped precursor were hand altered when necessary to ensure there were no glitches as the F2 component transitioned into the critical region.[2]

[2] We first attempted to use stimuli in which only the precursors were sinewave replicas, with the critical region being the same natural speech tokens used in Experiments 1 and 2. An LRE failed to emerge even in the original (unflipped) condition. We suspect this occurred because the extreme differences in signal characteristics and voice quality between the sinewave precursor and clear-speech target made it difficult to integrate the two.

Procedure

Prior to participating in the experiment, listeners were familiarized with sinewave speech to ensure they could perceive its phonetic content and hear the spoken words.

Familiarization session The familiarization session differed slightly across the two precursor conditions because of their differences in speech-likeness. In the original condition, familiarization consisted of listeners comparing alternating versions of the clear and sinewave precursors (described as computer speech, and presented at the spoken rate and slowed rates) so that they could hear their equivalence. This phase was self-paced and continued for as long as the participant desired. Next, each sinewave version was presented individually, after which the participant had to repeat the precursor aloud to the experimenter. Participants were permitted to listen to each precursor as many times as needed but did not move on to the next precursor until the phrase was spoken correctly. The experimenter provided feedback and corrected the participant when difficulty was encountered.

The familiarization session in the flipped condition was more elaborate because listeners needed exposure to the flipped precursors as well as sinewave speech. The session began as in the original condition but used four other phrases (nontarget items) from Experiment 1 instead of the target phrases in the current experiment. Once participants were able to hear the sinewave tokens as speech, they were introduced to flipped precursors, which now included the target stimuli and flipped versions of these four extra stimuli. Participants were not told these stimuli were derived from speech, only that past participants have commented that they sound like computer bleeps and whistles. Participants were informed that some of these would be presented during the test session and were allowed to listen to them as often as they liked.

The preceding two phases of familiarization were then combined to give participants in the flipped condition practice in hearing clearly the unflipped critical region after the flipped precursor. The four nontarget items were used, and repeated multiple times. This phase ended with listeners hearing the spoken-rate versions of the four flipped target items (precursors plus critical region) and reporting the words in the critical region. Performance on these items served as a test to assess listeners' ability to hear the words in the critical region, since the precursor was intelligible. Accuracy across participants averaged 83 %, and demonstrates that the familiarization session achieved its goal of ensuring that participants could hear most of the words in the critical region as intended and without prompting.[3]

[3] A less elaborate familiarization session, which included only exposure to sinewave replicas and the flipped precursors (not their combination or a test of comprehension of the critical region), yielded similar results. These data are available from the authors.

Test session The test session followed the procedure in Heffner et al. (2013). The experiment consisted of eight blocks of the eight sentences (half at each rate), for a total of 64 trials. No stimulus repeated until all had been presented in a block, and stimulus order was randomized within blocks. The start of each trial began with presentation of one of the stimuli. Immediately after its offset, two response alternatives appeared adjacent to each other in the middle row of the computer screen in front of the participant. One option was the noun plus the function word (*minor or*) and the other was the noun alone (*minor*). Participants were instructed to listen to the stimuli and press the button that corresponded to what was spoken. Response speed was not emphasized, and so reaction time data were not analyzed. Participants in the original condition were informed that the precursors heard in the familiarization session were the first part of longer phrases. They would now also hear the continuations of the phrases, and have to decide which alternative was spoken in the continuation. Participants in the flipped condition were told that flipped precursors would be followed by a few of the computer speech words, as in the final part of the practice session, only now the task would be to determine which of the two alternatives on the screen was spoken. The next trial began 1.5 s after a response. There was a 4-s timeout. Six practice trials preceded the test session, which lasted 10 minutes.

Results and Discussion

The data were scored and analyzed as in Experiment 1. The proportion of time listeners responded that the function word was spoken is plotted in Fig. 2 across the four conditions. The results provide additional evidence of the LRE being speech specific. There was a large (.22) effect of speech rate in the original condition but none (-.017) in the flipped condition. The results of a mixed-models analysis yielded only a reliable interaction of the two variables ($b = 1.205$, $p < .01$). Comparisons of the effect of rate within each precursor condition proved reliable only in the original condition ($b = -0.225$, $p < .001$; flipped: $b = 0.018$, $p < .79$).

The pattern in Fig. 2 held up when more fine-grained analyses were performed on the data. The majority of participants in the original condition (14 of 15 with one tie) showed the LRE in the prediction direction; only 6 of 14 (two ties) did so in the flipped condition. Similar consistency was obtained with the four items; all showed large (>.12) effects in the original condition. In the flipped condition, only one showed a .05 effect, with the other three showing small reversals. Last, the LRE was sizeable and fairly consistent across the course of the experiment. The rate effect was measured at four time points by aggregating data across pairs of adjacent blocks (e.g., 1 + 2, 3 + 4, etc.). In the flipped condition, the rate effect flip-flopped from slightly negative to slightly positive across quarters, beginning with a -.01 effect and ending with a -.07 effect; none of

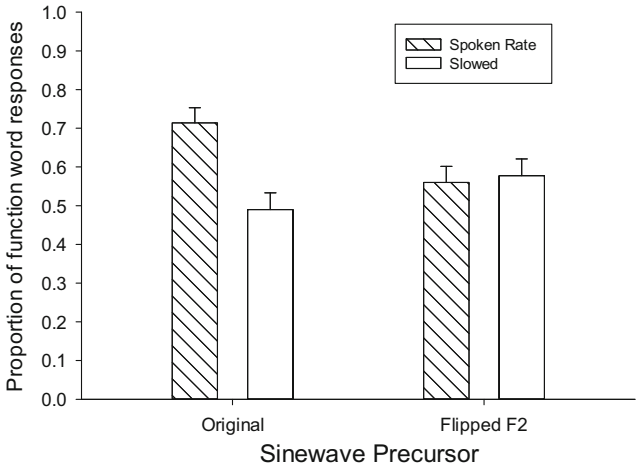

Fig. 2 Graph of mean function word reports as a function of speech rate in the original and flipped conditions of Experiment 3. Error bars represent 95 % confidence intervals

these differences was reliable. In the original condition, the rate effect was present starting in the first time point (.16) and increased to .23 by the last; all differences were reliable.

To further reinforce the conclusions of the current experiment, and connect them with those in Experiments 1 and 2, we reran the clear condition in Experiment 1 (22 participants) using sinewave replicas of all phrases that the authors considered at least moderately intelligible (38 of 48). If the rate effect depends on the intelligibility of the precursor, we should expect the size of the LRE to correlate positively with accuracy in understanding the precursor (scored as the number of syllables typed correctly). Other than a few practice trials, no familiarization with sinewave speech was provided so that we could observe the full range of individual differences in understanding it. Although the lack of familiarity with sinewave speech likely resulted in a small LRE overall (.09), the predicted correlation was sizeable and in the expected direction: Transcription accuracy of the precursor was positively correlated with the size of the LRE, $r(20) = .505$, $p < .017$, with LREs ranging from -.16 to .48 across individuals.

The results of Experiment 3 strengthen the claim that the LRE is driven by rate information in signals heard as intelligible speech. In contrast to Experiments 1 and 2, the precursors were virtually acoustically identical across conditions, yet they had distinctly different influences on perception, altering the frequency of function word reports only when heard as speech. The results also provide another demonstration that the LRE can be obtained using a classification task.

General Discussion

The present study probed the domain specificity of the LRE by examining the conditions necessary to produce it. Across three experiments, we asked whether rate information conveyed by multiple types of precursors (tones, clear speech, filtered speech, sinewave speech) were equivalent by measuring the ability of each to produce an LRE. The results are clear and consistent. Only stimuli perceived as intelligible speech produced an LRE. In Experiment 1, function word reports across the spoken-rate and slowed-rate conditions differed only when the precursor was clear speech. In Experiment 2, an effect of speaking rate was found only when the filtered precursors were disambiguated. In Experiment 3, sinewave precursors generated an LRE only when heard as speech. In contrast, in seven conditions (including replications) across Experiments 1 through 3, when the precursor was a tone sequence, primed or unprimed, or an altered spoken phrase rendered unintelligible by signal manipulation, an LRE was never found.

The results of Experiments 2 and 3 suggest that the acoustic properties of the precursor alone are not sufficient to induce the LRE; what also matters is how the precursor is processed. A speech precursor that was low-pass filtered produced no rate effect, but when disambiguated with a clear-speech prime so that participants could restore the words in the precursor at the specified rate, a healthy LRE emerged. The sinewave precursors in the original and flipped conditions of Experiment 3 were acoustically almost identical, yet only those heard as speech yielded an LRE. This selectivity of the LRE points to a degree of domain specificity in rate-based speech processing.

These findings contrast with those exploring rate-based effects in phonetic processing, which the evidence suggests are domain-general. The labeling boundary along a phonetic continuum can be altered by the rate of preceding speech (Miller 1981; Port 1979; Summerfield 1981). The location of the boundary can also be shifted by pulse trains of tones presented at different rates, as well as by sine tones that follow the amplitude contour of a short phrase (Gordon 1988). In these latter studies, the repeating tones occurred at regular intervals, and most listeners would agree they induce a sense of rate, fast or slow. In contrast, the results of Experiments 1 and 2 show that rate information conveyed by tone precursors was insufficient to generate the LRE.

Another reason why these two rate-based phenomena might require different explanations comes from a consideration of how context affects responding. The phonetic speech rate effect seems to serve the purpose of ensuring perceptual constancy and appears to be contrastive. As speech rate increases or decreases, the phonetic boundary shifts in a compensatory direction to ensure the phonetic percept remains stable. For example, at a fast rate of speech, the boundary on a /b/–/p/ continuum shifts to shorter values (more /p/ responses) so as to preserve the contrast for the listener. The LRE is a qualitatively different entrainment phenomenon and seems more akin to perceptual assimilation than contrast. The precursor establishes a given speech rate. The listener extrapolates this rate into the critical region, which causes perception of the critical region to assimilate to the rate of the precursor.

Wade and Holt (2005) suggested that rate-dependent processing may have multiple causes, some of which could be specific to speech or language. The current data support this contention, and raise the question of how to explain both phenomena within a single processing system. One possibility is that domain-specific processing occurs at a more abstract level of analysis. Sjerps et al. (2011) advanced this proposal in a study that addressed the question of domain-general processing in a different content area—that of contextual influences in vowel normalization. One dimension along which context varied was whether it was speech or nonspeech. Differences in responding as a function of these two stimulus types prompted Sjerps et al. (2011) to suggest that speech-specific effects might be occurring at a level of auditory analysis that is more abstract than that found with nonspeech contexts. The current data are in accord with this account. The LRE demonstrates a degree of processing selectivity, which may be indicative of a separate, higher level of analysis.

Mechanistically, we currently favor an explanation in which speech-rate effects arise from the engagement of endogenous neural oscillators that are attuned to the temporal periodicities and quasi-periodicities present in speech (Barbosa 2007; Byrd and Saltzman 2003; Cummins and Port 1998; Large and Jones 1999; McAuley and Jones 2003; Nam et al. 2006; Port 2003). Oscillatory activity in the brain occurs over a small range of frequencies (~1–40 Hz), which include the temporal frequencies of linguistic units (phonemes, syllables). This correspondence has led some researchers (Doelling et al. 2014; Giraud and Poeppel 2012) to propose that the initial neural coding of speech occurs from the coupling of oscillatory networks that separately synchronize to the slow (syllabic, theta band) and the fast (phonemic, gamma band) amplitude fluctuations in speech. Empirical evidence exploring this relationship continues to show a connection between the two. Cortical activity in the 4 to 8 Hz range will phase lock to attended speech (Doelling et al. 2014; Kerlin et al. 2009; Luo and Poeppel 2007), and the magnitude of the phase locking increases when listening to intelligible speech (Peelle et al. 2013). In addition, Goss et al (2013; Arnal et al. 2014) report evidence suggestive of hierarchical coupling between slow and fast oscillatory networks. Although speech encoding might also involve phase-locking to frequency fluctuations as well as amplitude fluctuations (Henry and Obleser 2012, 2013) these studies identify a neural mechanism that is associated with the temporal encoding of speech (i.e., entrainment) that may help explain the LRE.

The speech-specificity of the LRE might also reflect the engagement of other processes. Specifically, we conjecture that the LRE comes about due to the recruitment of processes involved in speech production. Knowledge of how speech is produced provides a means of arriving at a reinterpretation of the critical region that is phonetically coherent and plausible (aided by semantic and discourse knowledge). For a given rate of speech, knowledge of how the coordinative motor actions of the articulators will unfold over time could be used to estimate probable trajectories over the critical region, in a manner similar to analysis by synthesis (Halle and Stevens 1962; see also Hickok 2012). The listener's knowledge of her own productions (and plausible words) can greatly constrain possible interpretations of the ambiguous phonetic material in the critical region. Essentially, the knowledge of the intrinsic timing of the motor system is recruited in the service of perception, in the same way that phonological and lexical knowledge are used to guide phonetic and word perception. To be effective, such production constraints would have to be sensitive to the current rate of speech. An entrainment mechanism, such as the one described above, would serve this purpose. As an example, a possible internal production model that listeners might use for articulatory prediction could resemble the task-dynamic model of speech production (Saltzman et al. 2008), in which oscillators at different time scales (subsyllabic, syllabic, word, phrase) are coordinated to ensure smooth adjustments for the production of fast and slow speech.

In summary, the results of the present study support the hypothesis that the LRE is driven by a timing mechanism that requires hearing input (precursor) as intelligible speech. Across three experiments, precursors that were perceived as intelligible speech induced an LRE, in spite of the fact that those precursors differed greatly in their acoustic properties, even consisting of sine-wave tones. Precursors that were not heard as intelligible speech, even when consisting of filtered speech content, did not yield an LRE. These findings provide new information regarding the nature of the timing mechanism underlying speech perception, and are argued to reflect the operation of an oscillatory mechanism in which listeners entrain to speech rate under the guidance of higher-level linguistic content to ensure accurate uptake and interpretation of what a talker said.

Acknowledgments We thank Hartman Brawley, Karin Luk, Chloe Meyer, Jason Miao, Kristen Peters, and Sara Stanutz for assistance in testing participants and scoring data, and Chris Heffner for help in constructing stimuli. Correspondence concerning this article should be addressed to Mark Pitt, pitt.2@osu.edu.

References

Arnal, L. H., Doelling, K. B., & Poeppel, D. (2014). Delta–beta coupled oscillations underlie temporal prediction accuracy. *Cerebral Cortex*. doi:10.1093/cercor/bhu103

Baer, T., & Moore, B. C. J. (1993). Effects of spectral smearing on the intelligibility of sentences in noise. *Journal of the Acoustical Society of America, 94*(3), 1229–1241.

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

Barbosa, P. (2007). From syntax to acoustic duration: A dynamical model of speech rhythm production. *Speech Communication, 49,* 725–742.

Bates, D., Maechler, M., & Bolker, B. (2012). *lme4: Linear mixed-effects models using S4 classes* (R package version 0.999375-42). https://cran.r-project.org/web/packages/lme4/index.html

Best, C. T., Morrongiello, B., & Robson, R. (1981). Perceptual equivalence of acoustic cues in speech and nonspeech perception. *Perception & Psychophysics, 29*(3), 191–211. doi:10.3758/bf03207286

Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics, 31,* 149–180.

Cummins, F., & Port, R. F. (1998). Rhythmic constraints on stress timing in English. *Journal of Phonetics, 26,* 145–171.

d'Alessandro, C., & Mertens, P. (1995). Automatic pitch contour stylization using a model of tonal perception. *Computer Speech and Language, 9*(3), 257–288.

Davis, M. H., Johnsrude, I. S., Hervais-Adelman, A., Taylor, K., & McGettigan, C. (2005). Lexical information drives perceptual learning of distorted speech: Evidence from the comprehension of noise-vocoded sentences. *Journal of Experimental Psychology: General, 134*(2), 222. doi:10.1037/e537052012-126

Dilley, L. C., & Pitt, M. (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., Morrill, T., & Banzina, E. (2013). New tests of the distal speech rate effect: Examining cross-linguistic generalizability. *Frontiers in Language Sciences, 4*(1002), 1–13. doi:10.3389/fpsyg.2013.01002

Doelling, K., Arnal, L., Ghitza, O., & Poeppel, D. (2014). Acoustic landmarks drive delta–theta oscillations to enable speech comprehension by facilitating perceptual parsing. *NeuroImage, 85,* 761–768. doi:10.1016/j.neuroimage.2013.06.035

Drullman, R., Festen, J. M., & Plomp, R. (1994a). Effect of reducing slow temporal modulations on speech reception. *The Journal of the Acoustical Society of America, 95*(5), 2670. doi:10.1121/1.409836

Drullman, R., Festen, J. M., & Plomp, R. (1994b). Effect of temporal envelope smearing on speech reception. *Journal of the Acoustical Society of America, 95,* 1053–1064.

Elliott, T., & Theunissen, F. (2009). The modulation transfer function fro speech intelligibility. *PLOS Computational Biology, 5*(3), e1000302. doi:10.1371/journal.pcbi.1000302

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

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

Gordon, P. A. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Attention, Perception, & Psychophysics, 43,* 137–146.

Gross, J., Hoogenboom, N., Thut, G., Schyns, P., Panzeri, S., Belin, P., & Garrod, S. (2013). Speech rhythms and multiplexed oscillatory sensory coding in the human brain. *PLoS Biology, 11*(12), e1001752–e1001752. doi:10.1371/journal.pbio.1001752

Guenther, F. H., Espy-Wilson, C., Boyce, S., Matthies, M., Zandipour, M., & Perkell, J. S. (1999). Articulatory tradeoffs reduce acoustic variability during American English /r/ production. *Journal of the Acoustical Society of America, 105*(5), 2854–2865.

Halle, M., & Stevens, K. N. (1962). Speech recognition: A model and a program for research. *IEEE Transactions on Information Theory, 8*(2), 155–159. doi:10.1109/tit.1962.1057686

Heffner, C., Dilley, L. C., McAuley, J. D., & Pitt, M. (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

Henry, M. J., & Obleser, J. (2012). Frequency modulation entrains slow neural oscillations and optimizes human listening behavior. *Proceedings of the National Academy of Sciences, 109*(49), 20095–20100. doi:10.1073/pnas.1213390109

Henry, M. J., & Obleser, J. (2013). Dissociable neural response signatures for slow amplitude and frequency modulation in human auditory cortex. *PLOS ONE, 8*(10), e78758. doi:10.1371/journal.pone.0078758

Hickok, G. (2012). Computational neuroanatomy of speech production. *Nature Reviews Neuroscience*. doi:10.1038/nrn3158

Hillenbrand, J., Getty, L. A., Clark, M. J., & Wheeler, K. (1995). Acoustic characteristics of American English vowels. *Journal of the Acoustical Society of America, 97*(5), 3099–3111.

Kerlin, J., Shahin, A., & Miller, L. (2009). Gain control of cortical speech representations by selective attention in a "cocktail party.". *NeuroImage, 47,* S42. doi:10.1016/s1053-8119(09)70005-9

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*(1), 119–159. doi:10.1037/0033-295X.106.1.119

Luo, H., & Poeppel, D. (2007). Phase patterns of neuronal responses reliably discriminate speech in human auditory cortex. *Neuron, 54,* 1001–1010.

McAuley, J. D., & Jones, M. R. (2003). Modeling effects of rhythmic context on perceived duration: A comparison of interval and entrainment approaches to short-interval timing. *Journal of Experimental Psychology: Human Perception and Performance, 29*(6), 1102–1125. doi:10.1037/0096-1523.29.6.1102

Mehler, J., Jusczyk, P. W., Lambertz, G., Halsted, N., Bertoncini, J., & Amier-Tison, C. (1988). A precursor of language acquisition in young infants. *Cognition, 29,* 143–178.

Mertens, Piet (2004). The Prosogram : Semi-Automatic Transcription of Prosody based on a Tonal Perception Model. In B. Bel & I. Marlien (eds.) Proceedings of Speech Prosody 2004, Nara, Japan. p 23–26.

Miller, J.L. (1981). Phonetic perception: Evidence for context-dependent and context-independent processing. *Journal of the Acoustical Society of America, 69*(3), 822–831. doi:10.1121/1.385593

Morrill, T., Dilley, L., McAuley, J. D., & Pitt, M. (2014). Distal rhythm influences whether or not listeners hear a word in continuous speech: Support for a perceptual grouping hypothesis. *Cognition, 131*(1), 69–74. doi:10.1016/j.cognition.2013.12.006

Nam, H., Goldstein, L., & Saltzman, E. (2006). *Dynamical modeling of supragestural timing.* Paper presented at the Proceedings of the 10th Laboratory Phonology Conference, Paris, France.

Peelle, J. E., & Wingfield, A. (2005). Dissociations in perceptual learning revealed by adult age differences in adaptation to time-compressed speech. *Journal of Experimental Psychology: Human Perception and Performance, 31*(6), 1315–1330. doi:10.1037/0096-1523.31.6.1315

Peelle, J. E., Gross, J., & Davis, M. H. (2013). Phase-locked responses to speech in human auditory cortex are enhanced during comprehension. *Cerebral Cortex, 23*(6), 1378–1387. doi:10.1093/cercor/bhs118

Port, R. F. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7,* 45–56.

Port, R. F. (2003). Meter and speech. *Journal of Phonetics, 31,* 599–611.

Reinisch, E., Jesse, A., & McQueen, J. M. (2011a). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech, 54*(2), 147–165. doi:10.1177/0023830910397489

Reinisch, E., Jesse, A., & McQueen, J. M. (2011b). 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

Remez, R. E., Rubin, P. E., Pisoni, D. B., & Carrell, T. D. (1981). Speech perception without traditional speech cues. *Science, 212,* 947–949. doi:10.1126/science.7233191

Saltzman, E., Nam, H., Krivokapic, J., & Goldstein, L. (2008). A task-dynamic toolkit for modeling the effects of prosodic structure on articulation. *Proceedings of the Fourth International Conference on Speech Prosody* Campinas, Brazil, 175–184.

Shannon, R. V., Zeng, F. G., Kamath, V., Wygonski, J., & Ekelid, M. (1995). Speech recognition with primarily temporal cues. *Science, 270,* 303–304. doi:10.1126/science.270.5234.303

Sjerps, M. J., Mitterer, H., & McQueen, J. M. (2011). Listening to different speakers: On the time-course of perceptual compensation for vocal-tract characteristics. *Neuropsychologia, 49,* 3831–3846. doi:10.1016/j.neuropsychologia.2011.09.044

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

Turk, A. E., & Sawusch, J. R. (1997). The domain of accentual lengthening in. *American English Journal of Phonetics, 25*(1), 25–41. doi:10.1006/jpho.1996.0032

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics, 28,* 397–440.

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