Proceedings of Meetings on Acoustics

Volume 19, 2013 http://acousticalsociety.org/

ICA 2013 Montreal
Montreal, Canada
2 - 7 June 2013

Speech Communication
Session 5aSCb: Production and Perception II: The Speech Segment (Poster Session)

5aSCb51. Canadian oats and Canadian goats: Comparing distal cues to segmentation and segments

Christopher C. Heffner* and Rochelle S. Newman

***Corresponding author's address: Program in Neuroscience and Cognitive Science, University of Maryland, College Park, 1401 Marie Mount Hall, College Park, MD 20742, heffner@umd.edu**

Distal timing cues (specifically, the speech rate of sentences temporally removed from a point of ambiguity in speech) have been shown to weakly modulate segmental perception (i.e., the perception of basic speech sounds, like "p" and "b"; Newman & Sawusch, 1996) but strongly affect the perception of word boundaries (i.e., the perception of separation between words in fluent speech; Dilley & Pitt, 2010). However, no study as of yet has directly compared the two classes of percept using identical manipulations. In this study, we will examine the role of distal timing cues to segmental perception and word segmentation using the same distal contexts (e.g., "The merchant sold Canadian oats/notes", a word boundary distinction, or "The merchant sold Canadian coats/goats", a voicing-based segmental distinction). Distal speech rate will be artificially slowed to an identical extent for both types of contrast. We predict that categorical perception leads distal context effects to be much stronger on word segmentation ("...Canadian oats/notes") than on segmental perception ("...Canadian coats/goats"). This research may help bridge speech perception theories that have been developed for each class of percept, and further clarify the role of distal cues in speech perception.

Published by the Acoustical Society of America through the American Institute of Physics

 [DOI: 10.1121/1.4799792]
Received 22 Jan 2013; published 2 Jun 2013

INTRODUCTION

Given the amount of variability that exists in speech, human speech perception capabilities are remarkable. Consider the variability that exists in speech rate. The speech rate of fast-talking auctioneers is much more rapid than the measured cadences of old-school orators; similarly, a person talking on the phone with a sibling may talk at a different rate than that same person calling to report a fire. Yet, given enough exposure, the comprehension of fast and slow speech can approach similar levels.

Variation in speech rate can lead to further difficulties in speech perception related to a couple of the tasks listeners are faced with in understanding speech: the perception of basic sound contrasts and the perception of boundaries between words. The former problem is referred to as segmental perception, while the latter problem is known as word segmentation. In both cases, the speech signal may be highly ambiguous. Despite this, perception of segments and word boundaries is most often instantaneous, depending in both cases on the presence of multiple cues to segments and to word boundaries (Lehiste, 1960).

The similarities between the two phenomena are readily apparent. Yet not much research has been devoted to directly comparing the processes of segmental perception and word segmentation. This is particularly interesting in light of the ways the two percepts diverge from each other. One particularly interesting divergence emerges with respect to speech rate. Studies of the effects of compensation for the effects of speech rate on segmental perception are fairly common (see, e.g., Miller, 1981); depending somewhat on the study, the effects of speech rate seems to persist over only a short time window (Newman & Sawusch, 1996). Conversely, the effects of speech rate on word segmentation have only been recently examined, and have been shown to be maintained for a much longer duration, with what have been termed “distal” effects of speech rate on word segmentation persisting across an entire sentence (Dilley & McAuley, 2008; Dilley & Pitt, 2010).

Here, we examine the effects of distal speech rate on segmental perception and word segmentation, with an eye for the ways in which these two percepts may fruitfully be compared to each other.

Distal Effects on Segmental Perception

The fact that the duration of various sub-parts of segments can vary is well-established; however, this variation can come from multiple sources. Even in languages without canonical phonemic duration contrasts, small differences in duration can cue the difference between the /æ/ vowel of ‘bad’ and the /ɛ/ vowel of ‘bed’ (Hillenbrand, Getty, Clark, & Wheeler, 1995), the /ʃ/ of ‘ship’ versus the /tʃ/ of ‘chip’ (Repp, Liberman, Eccardt, & Pesetsky, 1978), and the /b/ of ‘bad’ versus the /p/ of ‘pad’ (Lisker & Abramson, 1964). Differences in duration can also signal changes in lexical stress, as in the contrast between ‘contract’ the noun and ‘contract’ the verb (Lieberman, 1960). Most importantly to this discussion, the durations of various acoustic cues to the identity of consonants and vowels are all affected by speech rate (Miller, 1981). The changes in consonantal duration as a result of speech rate vary from consonant to consonant, but, in general, faster speech rates have shorter-duration consonants (Crystal & House, 1982, 1988). Thus, the question of how perceived speech rate can lead to changes in segmental perception is intimately tied to broader questions of how listeners are able to make use of changes in duration more generally, especially how they are able to know which durational features are the result of speech rate versus other duration-related aspects of speech.

Some studies have found effects of speech rate (or a close proxy, adjacent segment duration) on segmental perception. Miller and Dexter (1988) found that listeners almost always use speech rate information in determining consonant voicing, though the specific information used varied as a function of response speed and task demands. The effects of adjacent consonant duration also influence which members of a category are perceived to be exemplars (Miller & Volaitis, 1989). Yet few of these studies specifically manipulate *distal* speech rate measures, as most of them (see, e.g., Repp, 1982 for an early review) focus on immediately adjacent consonants. Kidd (1989) provided one exception; in his study, Kidd found that speech rate effects on voice onset time (VOT) distinctions in consonants depended on distal speech rate patterns across an entire sentence.

Most previous studies that have looked at distal speech rate effects on segmental perception, by contrast, have found weak to no support for the idea that distal rate cues can have influence. Summerfield (1981), like Miller and colleagues, found that the duration of segments adjacent to a consonant ambiguous with respect to voicing could influence the perception of voicing contrasts. However, manipulating the duration of consonants more removed than direct adjacency had no influence on segmental contrasts. Many segmental contrasts—between /ʃ/ and/tʃ/, /t/ and /s/, /b/ and /p/, and /d/ and /t/—only seem to be affected by the rate cues encoded in the duration of adjacent consonants, but no further. A window of approximately 300-400 milliseconds, beyond which speech rate cannot directly influence segmental perception, has been proposed (Newman & Sawusch, 1996; Sawusch & Newman, 2000). Even if the influence of distal speech rate on segmental perception exists, it may be very small. Newman and Sawusch (2009) found that distal speech rate, even the speech rate of sentences played to the contralateral ear from an attended stimulus, can in fact significantly influence perception of voicing contrasts. But the effects were very small, with changes in speech rate leading to as low as 3-4% changes in the probability of hearing a /g/ sound (versus a /k/ sound) for the stimuli used.

Distal Effects on Word Segmentation

In contrast to findings in the segmental perception literature, the effects of speech rate on word segmentation are less established. In production, prosodic boundaries, such as the word boundary, are typically said to be associated with lengthening in segmental duration (Vaissière, 1983; Turk & Shattuck-Hufnagel, 2000), and the production of lengthening around prosodic boundaries shows up as early in development as 10 months (DePaolis, Vihman, & Kunnari, 2008). This prosodic lengthening might be conceived as modulation in speech rate around prosodic boundaries (Byrd & Saltzman, 2003). Presumably, this lengthening of prosodic boundaries could be used by listeners to segment the utterance. If so, durational changes caused by speech rate could also potentially influence how listeners segment an utterance. Understanding the ways that segmentation might be affected by modifications in duration that are the result of speech rate changes could prove particularly informative.

However, in perception, investigation of the role of speech rate, particularly distal speech rate, has begun only recently. Dilley and McAuley (2008) investigated the role of prosodically-conditioned expectancies, including patterns in speech rate, on word segmentation. To do this, they used syllabic sequences with ambiguous word segmentation in the last four syllables: for example, *chocolate lyric down town ship wreck*. The syllabic sequence *down town ship wreck* can be parsed as either *downtown shipwreck* or *down township wreck*. Participants were asked to transcribe the last word they heard in the sequence, and their responses were used to indicate how listeners segmented the sequence. For example, if listeners perceived a word boundary between the second and third syllables, they would have heard *downtown shipwreck* and reported *shipwreck* as the last word they heard. Crucially, participants heard sentences that varied in the speech rate (and/or fundamental frequency) information in the *first* four, unambiguous syllables, in this example *chocolate lyric*. Participants segmented the last four syllables in the sequence in line with the patterns established by the speech rate patterns of the first four syllables. For instance, increasing the duration of the syllable *down* to be in line with the length of the words *chocolate* and *lyric* led to listeners reporting the final word in the syllable string as *shipwreck*; if the duration of *down* was much shorter than the expectations created by *chocolate* and *lyric*, however, listeners more frequently heard *wreck*.

The research of Dilley and McAuley (2008) was later extended by Dilley and Pitt (2010) to more natural sentential contexts. In their study, Dilley and Pitt used sentence fragments with acoustically-reduced function words in them; this acoustic reduction led to acoustic ambiguity as to the existence of a word boundary. For example, the sentence fragment *Anyone must be a minor or child…* can be acoustically ambiguous to the word boundary signaling the existence of the word *or* if *or* is pronounced [ɚ], which is identical to the last segment of the preceding word *minor*. That is, listeners could potentially hear the speech fragment as *Anyone must be a minor child…* rather than *Anyone must be a minor or child….* In their study, the authors kept the acoustic information proximal to the possible word boundary (e.g., *nor or ch*) as originally recorded, but slowed down the distal speech rate in the rest of the sentence. Participants reported hearing function words much less often in distally-slowed versions of the sentence fragments without doing anything to the region immediately adjacent to the word boundary signaling the function word itself. To put this another way, if the listener believes a speaker is talking slowly, a production of [ɚ] that would have been perceived as the last syllable of *minor* and the function word *or* at a normal speech rate would instead be perceived as just the last syllable of *minor*. Similar results have since been obtained in Dutch (Reinisch, Jesse, & McQueen, 2011).

Comparison of Word Segmentation to Segmental Perception

Clearly, there are some stark differences between the findings of studies on the effects of distal speech rate on segmental perception versus results from studies on the effects of distal speech rate on word segmentation. In particular distal effects seem consistently strong in studies of word segmentation, but are often weak to non-existent in studies of segmental perception (with the caveat that there are stark disparities between the two domains in the number of studies conducted). However, the origin of these differences is not entirely clear.

Consider the methodological differences between the two types of studies, using Newman and Sawusch (2009) and Dilley and Pitt (2010) as exemplars of either class of studies. In Newman and Sawusch (2009), listeners heard variations on a single sentential context and had to rank the last word they heard in the sentence as anything from a "good, clear k" to a "good, clear g" on a six-point Likert scale. As such, there was no variation in the amount of distal context across trials. Moreover, listeners were exposed to the same sentence frame repeatedly and likely were tempted to ignore that context (out of boredom, if for no other reason). In Dilley and Pitt (2010), listeners had to listen to 120 different sentence fragments (including 70 filler fragments) and transcribe the full sentence fragment after it was played. Amount of distal context varied widely across trials. Newman and Sawusch (2009) used two talkers; Dilley and Pitt (2010) used twelve. In Newman and Sawusch (2009), distal speech rate was modulated by speaker productions at slow to fast rates; in Dilley and Pitt (2010), by contrast, distal speech rate was altered using the Pitch-Synchronous Overlap and Add (PSOLA) technique, which causes uniform changes across the speech being manipulated. It is not clear that natural and artificial slowing are truly identical; in fact, given that consonantal durations are rate-modulated to a different extent depending on the consonant (Crystal & House, 1982, 1988), such a conclusion is unlikely.

These differences are understandable, given the differences in labs and research methods for each study. However, they prevent a direct comparison between the two types of study. If variation in sentential context, talker characteristics, or any of the other characteristics of the studies that differed is what led to the divergences in the strength of distal speech rate, this could provide interesting insights into methodology. However, more interesting would be an explanation for the difference in effect sizes that stems from characteristics of segmental perception versus those of word segmentation themselves. As such, in the experiment below, we propose an experiment directly comparing the effects of distal speech rate on segmental perception to word segmentation. We use Newman and Sawusch (2009) and Dilley and Pitt (2010) as models and present preliminary production data for the types of contrast. This should allow us to examine the reasons why differences in the strength of distal timing effects on speech perception have been found.

METHODS

Participants

Talkers

Talkers for this experiment were recruited from the University of Maryland, College Park language science community. We recruited both male and female talkers, all native monolingual speakers of American English. Talkers were recorded using a Shure SM81 microphone. Recording sessions were programmed in PsychoPy experimental suite (Peirce, 2007). Talkers were instructed to speak quickly and accurately, but were encouraged to speak informally. Recording trials were paced so as to encourage talkers to speak relatively rapidly; this caused some trials to be cut off on recordings. These trials were discarded for further analysis, as were trials with speech errors. Talkers made the recordings over one to two sessions, lasting between a half an hour to an hour.

TABLE 1. Characteristics of talkers used in this study so far.

Talker	Gender	Age
1	F	17
2	F	23
3	M	23

Listeners

We will recruit approximately 60 listeners from participant populations at the University of Maryland, College Park. All listeners will be native monolingual American English speakers with no history of speech or language disorder or hearing impairment.

Materials

Stimuli

Stimuli were constructed to maximize similarity between experimental items across conditions. The most important manipulations were in the critical region, the last two words of each stimulus. Here, stimuli could vary by ambiguity type: segmental or lexical. Segmentally-ambiguous critical regions contained critical consonants that were ambiguous in the voicing qualities of either a word-initial or word-final stop consonant. For example, they could be perceived as either voiceless /p/ or voiced /b/. Lexically-ambiguous critical regions, on the other hand, contained critical consonants that were ambiguous to the presence of a word-initial or word-final continuant consonant identical to an adjacent continuant. That is, they were ambiguous as to where a word boundary occurred. Pairs of semantically-ambiguous critical regions and lexically-ambiguous critical regions were matched up to form stimulus sets. Wherever possible, stimulus sets differed minimally in segmental content. For example, the segmentally-ambiguous pair "Canadian goats" and "Canadian coats" was matched up with the lexically-ambiguous pair "Canadian oats" and "Canadian notes". Such similarity was not possible in all cases; as such, pairs were matched to form sets that could more easily take identical distal contexts.

A second manipulation was of distal context type; stimuli could have either an artificial or a natural distal context. Natural contexts were 4 to 7 syllables in length and formed a complete sentence along with a set of four critical regions; artificial contexts were a standard carrier phrase ("Jessica said…") with the possible addition of adverbs (i.e., "Jessica once said…", "Jessica often said…", "Jessica frequently said…") to match the number of syllables found in the distal context for equivalent natural contexts. In the perception experiment to come, listeners will only hear either natural or artificial contexts. Participants in the natural context condition will hear a wide variety of sentences with many possible contexts. Artificial context participants, meanwhile, will only hear variants of the carrier phrase. See the description of the experiment in "Design and Procedure" below. Examples of experimental stimuli are given below for a complete set ("Canadian c/g/n/otes").

TABLE 2. An example of a stimulus set used in this study.

Context	Ambiguity	Pair Member	Distal material	Critical Region
Natural	Segmental	Voiceless	The merchant sold	Canadian coats.
Natural	Segmental	Voiced	The merchant sold	Canadian goats.
Natural	Lexical	Present	The merchant sold	Canadian notes.
Natural	Lexical	Absent	The merchant sold	Canadian oats.
Artificial	Segmental	Voiceless	Jessica said	"Canadian coats."
Artificial	Segmental	Voiced	Jessica said	"Canadian goats."
Artificial	Lexical	Present	Jessica said	"Canadian notes."
Artificial	Lexical	Absent	Jessica said	"Canadian oats."

Manipulations

Sentences were divided into two portions for purposes of manipulation: a proximal region, and a distal region. The cutoff between the two regions was set at the next earliest syllable onset to a point 400 milliseconds before the onset of the acoustically-ambiguous consonants within the critical regions (e.g., 400ms before the onset of /k/ or /g/ in 'coats' and 'goats' or the second /n/ in 'Canadian' in 'Canadian notes' and 'Canadian oats' for the examples given in Table 2). The proximal region was defined as anything after that cutoff; the distal region was defined as anything before that point. The 400ms window was chosen based off the proposal in Newman and Sawusch (1996) of a 400 ms time window over which distal effects may be exerted on segmental contrasts. Talker 1, for example, initiated an /n/ at approximately 1.7 seconds into her recording of the artificial version of "Canadian oats". 400ms before this point is in the middle of the /ej/ in 'Canadian'. The closest previous syllable onset to /ej/ is the first /n/ in the critical region; as such, for this speaker and this recording, the distal region was defined as anything in the first syllable of "Canadian" and before, while the proximal region was defined as everything after and including the second syllable of the word "Canadian".

For each version of each sentence, various measurements were taken to assess the length of certain segments that were of critical importance for determining the percept in the signal. These measurements, collectively, will yield a "critical length" to be exploited in creating maximally ambiguous forms of each stimulus. For lexically-ambiguous stimuli, the length of the continuant consonant (e.g., /n/ in 'Canadian oats'/'Canadian notes') was measured. The length of the continuant consonant represents the primary cue for the existence of a second consonant at the word boundary (see, e.g., Shatzman & McQueen, 2006). For segmentally-ambiguous stimuli, meanwhile, the measurement taken depended on whether the contrast was word-initial or word-final. Word-initial consonants were measured for voice-onset time (VOT), the primary driver of the distinction between voiced and voiceless stimuli in initial position (Lisker and Abramson, 1964). Word-final consonants, meanwhile, were evaluated for the length of the immediately-preceding segment, as the length of the previously-adjacent segment is the primary cue for word-final voiced consonant contrasts (Raphael, 1972). Additionally, the length of the proximal and distal context was also assessed.

Preliminary duration measures are reported in Table 3 for a single talker. Raw values for the critical length of each stimulus pair were used to generate "critical ratios" that represented the duration differences between present and absent members of the pair (for lexically-ambiguous stimuli) and voiceless and voiced members of each pair (for word-final segmentally-ambiguous stimuli). Word-initial segmentally-ambiguous stimuli were treated somewhat differently. See later in this section for more details.

TABLE 3. Characteristics of the stimuli produced by Talker 1 in this study.

Pair Member	Distal Length	Proximal Length	Critical Length	Critical Ratio
Voiceless	0.990	1.256	0.163	0.96
Voiced	1.070	1.177	0.163	
Present	1.058	1.180	0.191	2.39
Absent	1.052	1.168	0.082	

The values computed above were then used to generate new versions of each stimulus. These new versions were intended to create a maximally-ambiguous critical region to either the voicing of the critical consonant (for segmentally-ambiguous stimuli) or the number of critical consonants (for lexically-ambiguous stimuli). For lexically-ambiguous stimuli, the duration of sonorant-present critical consonants was multiplied by the inverse of the average of the value listed in the critical ratio column and 1 to create new versions of the stimuli that, ideally, would be perceived as intermediate between a sonorant-present and sonorant-absent version of the stimuli when averaged across multiple items and participants. This may be illustrated most clearly with an example. If the average critical ratio were 2 (that is, the length of sonorant-present stimuli was, on average, twice the length of the sonorant-absent stimuli), the average of 2 and 1 would be 1.5. The inverse of 1.5 is 0.67. Thus, the duration of each sonorant-present stimulus, when multiplied 0.67, would yield a duration for that stimulus approximately midway between a sonorant-present and sonorant-absent stimulus.

A similar procedure was followed to create maximally-ambiguous versions of word-final segmentally-ambiguous stimuli by averaging across ratios between lengths for immediately-previous segments in originally-recoded stimuli and then multiplying the duration of immediately-preceding segments by a factor of one over the critical ratio in Table 3 above. Word-initial segmentally-ambiguous stimuli, meanwhile, were manipulated on an absolute scale. The VOT of word-initial voiceless segments was set to 35ms, a value that, based on previous literature (Lisker & Abramson, 1964), would be maximally ambiguous for perception. Further, for each maximally-ambiguous stimulus, two stimulus versions will be created: a normal distal-rate version, where distal context duration will be kept as originally recorded, and a slowed distal-rate version, where distal speech rate will be slowed by multiplying the duration of the distal portion by a fixed factor of 1.5 (i.e., increasing its duration to 150% of its initial value).

Design and Procedure

Two versions will exist of this experiment. In the "natural" version, participants will hear only natural context stimuli; in the "artificial" version, participants will hear only artificial context stimuli. This will allow for comparison between these two common methodologies in the distal speech rate literature. For both versions, participants will hear 48 experimental items. Half of the experimental items will be slowed-distal-context versions, while half will be original-context versions. Of the 24 experimental items played at each rate, 12 will be segmentally-ambiguous stimuli (6 word-initial, 6 word-final), while 12 will be lexically-ambiguous stimuli (6 word-initial, 6 word-final).

In the "natural" design, participants will hear approximately 52 filler items that will be randomly interspersed with the experimental items alongside the "natural" context stimuli they will hear. Of the filler items, half will be presented as originally recorded, while the other half will be presented while digitally slowed, again by a fixed factor of 1.5. In the “artificial” design, meanwhile, participants will hear no filler items, mirroring most studies that use critical phrases embedded in carrier sentences; they will only hear “artificial” context stimuli.

Participants will be asked to listen to each stimulus and write down what they hear, either the entire sentence (for natural context stimuli) or just the last two words (for the artificial context stimuli). Those transcriptions will then be used to assess the rate at which participants perceived the critical region as either sonorant-present or sonorant-absent (for lexically-ambiguous stimuli) or voiceless or voiced (for segmentally-ambiguous stimuli).

HYPOTHESES

We predict that there will be no differences in the strength of distal effects as a result of context type (i.e., natural vs. artificial); speakers will be equally affected by distal speech rate regardless of whether the context comes from a variant of a single, repeated carrier phrase or a more naturalistic sentence. However, we do predict that participants will be differently affected by distal rate manipulations depending on whether the ambiguity in question is segmental or lexical, leading to an interaction between ambiguity type and distal speech rate. For lexically-ambiguous contexts, we predict that slowing distal speech rate will result in a substantial decrease in perception of the critical consonant as present (e.g., “Canadian notes”) versus absent (e.g., “Canadian oats”). For segmentally-ambiguous contexts, we predict little to no effect of distal speech rate context on perception of the critical consonant as voiceless or voiced. If any effect emerges, in line with Newman and Sawusch (2009), we predict that participants will be more likely to hear the critical consonant as voiced.

CONCLUSIONS

Though word segmentation and segmental perception are often discussed by different authors in different literatures, research in both fields has sometimes shown a striking tendency to converge on similar questions. The topic of whether distal speech rate context can affect each process is an example of such convergence. In the segmental perception literature, the answer seems to be an only slightly-qualified negative (Newman & Sawusch, 1996, 2009); in the word segmentation literature, meanwhile, the answer has as yet been affirmative (Dilley & McAuley, 2008; Dilley & Pitt, 2010). However, stark methodological differences emerge when considering the conclusions of both tasks. Initial results in production indicate that differences in word segmentation and segmental properties in production can be found across talkers; we aim to extend these results to perception. In the end, we hope to start to determine whether the differences between the conclusions reached in each field are the result of the existing methodologies used or a more-fundamental difference between segmental perception and word segmentation.

ACKNOWLEDGMENTS

We would like to thank the members of UMD’s Language Development Lab for their help in running participants for this project, the talkers who recorded stimuli for their willingness to volunteer their time, and Bill Idsardi and the other members of the Brain and Language Interest Group (BLING) for their helpful discussion of methodologies.

REFERENCES

Byrd, D., and Saltzman, E. **(2003)**. “The elastic phrase: modeling the dynamics of boundary-adjacent lengthening.” J. Phon. **31**, 149-180.

Crystal, T. H., and House, A. S. **(1982)**. “Segmental durations in connected speech signals: Preliminary results.” J. Acoust. Soc. Am. **72**, 705-716.

Crystal, T. H., and House, A. S. **(1988)**. “Segmental durations in connected-speech signals: Current results.” J. Acoust. Soc. Am. **83**, 1553-1573.

DePaolis, R. A., Vihman, M. M., and Kunnari, S. **(2008)**. “Prosody in production at the onset of word use: A cross-linguistic study.” J. Phon. **36**, 406-422.

Dilley, L. C., and McAuley, J. D. **(2008)**. “Distal prosodic context affects word segmentation and lexical processing.” J. Mem. and Lang. **59**, 294-311.

Dilley, L. C., and Pitt, M. A. **(2010)**. “Altering context speech rate can cause words to appear or disappear.” Psych. Sci. **21**, 1664-1670.

Kidd, G. R. **(1989)**. “Articulatory-rate context effects in phonemic identification.” J. Exp. Psych.: Human Perc. and Perf. **15**, 736-748.

Hillenbrand, J., Getty, L. A., Clark, M. J., and Wheeler, K. **(1995)**. “Acoustic characteristics of American English vowels.” J. Acoust. Soc. Am. **97**, 3099-3111.

Lehiste, I. **(1960)**. “An acoustic-phonetic study of internal open juncture.” Phonetica **5**, S5-S54.

Lieberman, P. **(1960)**. “Some acoustic correlates of word stress in American English.” J. Acoust. Soc. Am. **32**, 451-454.

Lisker, L., and Abramson, A. S. **(1964)**. “Some effects of context on voice onset time in English stops.” Lang. & Speech. **10**, 1-28.

Miller, J. L. **(1981)**. *Perspectives on the study of speech* (Erlbaum, Hillsdale, NJ), Chap. 2, pp. 39-74.

Miller, J. L., and Dexter, E. R. **(1988)**. “Effects of speaking rate and lexical status on phonetic perception.” JEPHPP **14**, 369-378.

Miller, J. L., and Volaitis, L. E. **(1989)**. “Effect of speaking rate on the perceptual structure of a phonetic category.” Perc. & Psychophys. **46**, 505-512.

Newman, R. S., and Sawusch, J. R. **(1996)**. “Perceptual normalization for speaking rate: Effects of temporal distance.” Perc. & Psychophys. **58**, 540-560.

Newman, R. S., and Sawusch, J. R. **(2009)**. “Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another.” J. Phon. **37**, 46-65.

Turk, A. E., and Shattuck-Hufnagel, S. **(2000)**. “Word-boundary-related duration patterns in English.” J. Phon. **28**, 397-440.

Reinisch, E., Jesse, A., and McQueen, J. M. **(2011)**. “Speaking rate from proximal and distal contexts is used during word segmentation.” J. Exp. Psych.: Human Perc. and Perf. **37**, 978-996.

Raphael, L. J. **(1972)**. “Preceding vowel duration as a cue to the perception of the voicing characteristic of word-final consonants in American English.” J. Acoust. Soc. Am. **51**, 1296-1303.

Repp, B. H. **(1982)**. “Phonetic trading relations and context effects: New experimental evidence for a speech mode of perception.” Psych. Bull. **92**, 81-110.

Repp, B. H., Liberman, A. M., Eccardt, T., and Pesetsky, D. **(1978)**. “Perceptual integration of acoustic cues for stop, fricative, and affricate manner.” JEPHPP **4**, 621-637.

Sawusch, J. R., and Newman, R. S. **(2000)**. “Perceptual normalization for speaking rate II: Effects of signal discontinuities.” Perc. & Psychophys. **62**, 285-300.

Shatzman, K. B., and McQueen, J. M. **(2006)**. “Segment duration as a cue to word boundaries in spoken-word recognition.” Perc. & Psychophys. **68**, 1-16.

Summerfield, Q. **(1981)**. “Articulatory rate and perceptual constancy in phonetic perception.” J. Exp. Psych.: Human Perc. and Perf. **7**, 1074-1095.

Vaissière, J. **(1983)**. *Prosody: Models and measurements* (Springer, Berlin), Chap. 7, pp. 53-65.