Journal of Phonetics 46 (2014) 68–85

Contents lists available at ScienceDirect

Journal of Phonetics

journal homepage: www.elsevier.com/locate/phonetics

Prosodic patterning in distal speech context: Effects of list intonation and f0 downtrend on perception of proximal prosodic structure

Tuuli H. Morrill [a,b,*], Laura C. Dilley [a,b,c], J. Devin McAuley [b]

[a] Department of Communicative Sciences and Disorders, Michigan State University, USA
[b] Department of Psychology, Michigan State University, USA
[c] Department of Linguistics and Germanic, Slavic, Asian and African Languages, Michigan State University, USA

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form
12 March 2014
Accepted 16 June 2014
Available online 23 July 2014

Keywords:
Distal prosody
Word segmentation
Perceptual grouping
Intonation

ABSTRACT

Prosodic structure is often perceived as exhibiting regularities in the patterning of tone sequences or stressed syllables. Recently, prosodic regularities in the distal (non-local) context have been shown to influence the perceived prosodic constituency of syllables. Three experiments tested the nature of distal prosodic patterns influencing perceptions of prosodic structure, using eight-syllable items ending in ambiguous lexical structures (e.g., *tie murder bee*, *timer derby*). For distinct combinations of distal fundamental frequency (f0) and/or timing cues, two patterns were resynthesized on the initial five syllables of experimental items; these were predicted to favor prosodic grouping of final syllables such that listeners would hear a final disyllabic or monosyllabic word, respectively. Results showed distal prosodic patterning affected perceived prosodic constituency when (1) patterns consisted of regularity in timing cues, f0 cues, or both (Experiments 1–2); (2) items ended with either a low–high (Experiment 1) or a high–low (Experiment 2) tonal pattern; and (3) tonal patterns consisted of alternating low and high-pitched syllables with progressive f0 decrease, i.e., a 'downtrend' (Experiment 3). The results reveal that a variety of prosodic patterns in the distal context can influence perceived prosodic constituency and thus lexical processing, and provide a perceptually-motivated explanation for the organization of acoustic speech input into prosodic constituents.

1. Introduction

Intonation and duration patterns play important roles in signaling the prosodic structure of an utterance, and are essential components of natural, fluent speech (e.g., Shattuck-Hufnagel & Turk, 1996). The perception of prosodic structure facilitates language comprehension by providing cues to the syntactic structure of an utterance and aiding in lexical access (Carlson, Frazier, & Clifton, 2009; Gee & Grosjean, 1983; Pynte & Prieur, 1996; van den Berg, Gussenhoven, & Rietveld, 1992; Wagner, 2010; Watson & Gibson, 2004) and by providing indications of the location of word boundaries for segmenting the continuous speech stream (e.g., Cutler, 1990; Mattys & Melhorn, 2007). Certain word-level prosodic cues can be used in speech segmentation, such as word-initial stress in English (e.g., Christophe, Gout, Peperkamp, & Morgan, 2003; Cutler & Butterfield, 1992). In addition, local cues signaling the edges of prosodic phrases, which organize words into structural constituents according to syntactic, semantic, and pragmatic information, can also function as cues to word boundaries (Cutler, Dahan, & van Donselaar, 1997; Nespor & Vogel, 2007; Selkirk, 1984). For example, edges of prosodic domains tend to be marked with articulatory strengthening or phoneme lengthening (Cho, McQueen, & Cox, 2007; Fougeron & Keating, 1997; Turk & Shattuck-Hufnagel, 2000). Previous work has shown that the detection of prosodic boundaries directly at a locus of a structural ambiguity can influence the processing of lexical or semantic information (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Millotte, René, Wales, & Christophe, 2008; Salverda, Dahan, & McQueen, 2003), and can help to resolve syntactically ambiguous utterances by providing cues to the hierarchical organization of the spoken phrases (Carlson et al., 2009; Gee & Grosjean, 1983; Larkey, 1983; Pynte & Prieur, 1996; Truckenbrodt, 2001; van den Berg et al., 1992; Wagner, 2010; Watson & Gibson, 2004). Thus, the majority of previous work on the perception of prosodic constituency has

* Corresponding author at: Department of Communicative Sciences and Disorders, 1026 Red Cedar Road, Rm. B9 Oyer Speech and Hearing, Michigan State University, East Lansing, MI 48824, USA. Tel.: +1 517 432 7042.
E-mail address: tmorrill@msu.edu (T.H. Morrill).

0095-4470/$ - see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.wocn.2014.06.001

focused on acoustic information directly adjacent to a prosodic boundary. The current study focuses on effects of prosodic context *prior to* a locus of a lexical ambiguity and shows that, depending on the context, the same acoustic speech material can be interpreted with different prosodic structures, thereby influencing lexical processing.

Prosodic structure is generally described as consisting of a variety of domains which are organized into cumulatively larger domains, in what is referred to as the Prosodic Hierarchy (see Shattuck-Hufnagel & Turk, 1996, for a review). The domains of this hierarchy often exhibit repeating units marked by regularities in pitch, duration, and/or amplitude, and these correlates of linguistic rhythm show patterning in both speakers' productions and in the perception of speech (Couper-Kuhlen, 1993; Crystal, 1969; Dainora, 2001; Pierrehumbert, 2000). For example, listeners tend to hear stressed syllables as occurring at regular intervals, i.e., as perceptually isochronous (e.g., Lehiste, 1977). In addition, Pierrehumbert (2000) notes that the full inventory of theoretically possible combinations of pitch accents is never realized, but instead, the same pattern of accents is often repeated within a phrase. Dainora (2001) examined a large speech corpus to show that specific pitch accents frequently co-occur in predictable ways, also suggesting widespread regularities in prosodic patterning. Attempts to provide acoustic measures of isochrony have been mixed (Lehiste, 1977; Ohala, 1975) and there is evidence that planned utterances exhibit a greater degree of rhythmicity than unplanned utterances (Tilsen, 2012; Wheeldon & Lahiri, 1997). However, repeated intonation patterns and phrasal boundary types are particularly common in lists of items (Beckman & Ayers Elam, 1997; Schubiger, 1958), and occur in coordinate syntactic constructions of various types (Wagner, 2010). Such repetitions of accent patterns are widely referenced in research on a variety of languages, including German, Bengali, Japanese, Spanish, Italian, Korean, French, and English (e.g., Beckman & Pierrehumbert, 1986; D'Imperio, 2000; Grice, 1995; Hayes & Lahiri, 1991; Jun, 1993; Kim, 2004; Prieto, van Santen, & Hirschberg, 1995; Welby, 2003). Thus, the phenomenon of repeated intonation patterns appears to be common in spoken language. When such stretches of repeated patterning occur, they may have communicative value in generating prosodic expectancies for listeners. However, many questions remain as to the effects of these prosodic expectancies on linguistic processing.

Dilley and McAuley (2008) provided initial evidence that expectations about prosodic structure generated by distal (i.e., non-local) contextual prosodic regularities can influence subsequent word segmentation. In this work, stimuli consisted of auditory sequences beginning with four syllables comprising two trochaic words (e.g., *channel dizzy*) and ending with four syllables that could form compound words in more than one way (e.g., *footnote#bookworm, foot#notebook#worm,* etc.). Because the same phonological material could be interpreted in multiple ways, listeners had to use available cues to posit a prosodic structure for the final two syllables, which could be either separated by a prosodic boundary of a given level (e.g., prosodic word boundary), or could be part of the same prosodic unit at that level. Prosodic patterns were imposed on the initial five syllables of the sequence (e.g., *chan-nel-diz-zy-foot*) using f0 and duration cues both independently and simultaneously, in different conditions, to create distal contexts conducive to perceptual grouping of the final two syllables into one of two possible parses – either a disyllabic final word (e.g., *bookworm* in the "Disyllabic context") or with the two final syllables being separated by a word boundary, yielding a monosyllabic final word (e.g., *worm* in the "Monosyllabic context"). The acoustic characteristics of the final three 'proximal' syllables (e.g., *note-book-worm*) were always held constant, indicating that any effects of prosodic structure on perception would have to have originated from the distal context. To assess this, participants were asked to provide a free report of the final word and the proportion of disyllabic responses was recorded.

Consistent with an effect of distal prosody on lexical perception, participants in the Dilley and McAuley study reported more disyllabic final words with a Disyllabic distal context than with a Monosyllabic distal context. In addition, effects of prosodic expectations were greatest in the condition in which two prosodic cues signaled the expected grouping. Converging support for the distal prosody effect was found when a surprise visual word recognition test was used instead of a free word report task (Dilley & McAuley, Experiment 3); participants better remembered hearing a disyllabic word when it was previously heard with a congruent distal prosodic context than with an incongruent context.

As an overarching theoretical explanation of the effect of distal prosody on lexical perception, Dilley and McAuley (2008) proposed a *perceptual grouping hypothesis* that was motivated from work in non-speech auditory perception. A well-established finding in auditory perception is that patterns of tones varying in frequency, duration and amplitude induce periodic expectations about the grouping and accentuation of later sequence elements (Boltz, 1993; Jones, 1976; Jones & Boltz, 1989; McAuley, 2010; Thomassen, 1982). Performance in perceptual monitoring tasks is facilitated by rhythmic regularity, with increased accuracy for the detection of pitch, timbre, or time changes (Jones, Boltz, & Kidd, 1982; Jones, Moynihan, MacKenzie, & Puente, 2002; McAuley & Jones, 2003); for example, listeners more quickly and accurately detect deviations in a melodic sequence when the deviation occurs at an "expected" time point based on the rhythmic structure (temporal and pitch characteristics) of the preceding melody (Boltz, 1993). In the speech domain, the perceptual grouping hypothesis predicts that pattern of syllable groupings at the beginning of the utterance should carry over to how listeners tend to group syllables at the end of the utterance (leading to either disyllabic or monosyllabic parses of the final syllables depending on the expectations generated by the distal context).

Subsequent studies have shown that prosodic patterning in the distal context also has robust effects on the perception of prosodic structure when tested with distinct lexical forms. Dilley, Mattys, and Vinke (2010) used experimental items consisting of syllable strings in which endings of items which were lexically ambiguous and consisted of non-compound words (e.g., crisis#turnip vs. cry#sister#nip), contrasting with the compound word items of Dilley and McAuley (2008). The effect of distal prosodic patterning on the formation of expectations about syllable parsings was demonstrated in several experiments by Dilley et al. (2010), including their Experiment 1c, which used a lexical decision task in the context of a cross-modal phonological priming paradigm. The results using this paradigm demonstrated that distal prosody affects the speed and accuracy of lexical decision, and that these effects occur quite early in lexical processing, as opposed to being due to late-occurring, meta-linguistic strategies. The effect of distal prosody in online lexical processing was also illustrated by Brown, Salverda, Dilley, and Tanenhaus (2011), who used an eye-tracking paradigm to examine perception of a syllable which was ambiguous as to whether it constituted a single monosyllabic word or was part of

a disyllabic word (such as *pan*, as a monosyllabic word vs. *panda*, in which *pan-* is part of a disyllabic word) within the context of a grammatical utterance.

Although these previous studies convincingly demonstrate the effects of distal prosody on the segmentation of lexically ambiguous material, it is not clear to what extent the mechanisms involved in this processing can flexibly adapt to variation in the available prosodic cues. Because the phonetic correlates of prosody vary in naturalistic speech environments, it is important to test the extent to which distal prosodic effects can be elicited with not only different lexical materials, but distinct acoustic cues to prosodic constituency (e.g., f0 and duration) and with distinct intonation patterns. The current investigation tests these questions with a series of three experiments. The aim of Experiment 1 was to replicate the finding that distinct phonetic realizations of distal prosodic patterning can elicit effects on lexical perception with a different stimulus set (consisting of non-compound words), and to compare the effects of f0 and duration cues alone to the effects of both cues combined. The internal structure of compounds and the effects of this morphological structure on lexical access are often debated (for a review, see Fiorentino & Poeppel, 2007). Because of the possibility that each element of the compound is first accessed as a single morpheme, the string of monosyllabic compound elements in these stimuli could have potentially been accessed in a similar way to a series of monosyllabic words, with possible effects on the likelihood of interpreting the final syllable as monosyllabic. Though this is unlikely to have contributed to the observed effect of different distal prosodic patterns on word segmentation (see Dilley and McAuley (2008) for discussion), it is important to establish that the effects of distal prosody on segmentation can occur across a variety of morphological structures. Both Dilley et al. (2010) and Brown et al. (2011) used a combination of f0 and duration cues; combined cues had been shown in the original work of Dilley and McAuley (2008) to yield the strongest effects of distal prosodic patterning. However, independent effects of f0 and duration cues have so far been shown only in the original investigation of distal prosody, where morphologically atypical words were used, i.e., compound items (Dilley & McAuley, 2008).

Experiment 2 tested whether distal prosodic effects would occur with a distinct intonation pattern, namely a final (HL) sequence, replacing every H tone in Experiment 1 with an L tone and vice versa (see Fig. 3). The repeating LH and HL tonal patterns used in Experiments 1 and 2 respectively are both typical list intonation contours of English (Ashby, 1978; Beckman & Ayers Elam, 1997; Schubiger, 1958). The phonetic realization of intonation patterns in English can result in either high or low tonal elements being realized on a given syllable (e.g., Beckman & Pierrehumbert, 1986); the perceptual grouping hypothesis (Dilley & McAuley, 2008) would predict that perception of prosodic constituency should also accommodate varied realizations of intonation contours. However, previous studies have all used the same intonation pattern on the final three syllables of experimental stimuli: a High tone on the antepenultimate syllable, followed by a Low–High tone sequence on the penultimate and final syllables, respectively.

Finally, Experiment 3 tested if the distal prosody effect could be elicited using downtrend patterns. In order for a perceptual grouping hypothesis to make predictions which apply to natural speech more generally, the effects of distal prosody should be elicited even when the alternation of High and Low tones is abstract and the absolute frequencies are not maintained. So far, all experiments using distal prosodic manipulations have used an identical repetition of the same absolute f0 levels across the entire utterance. Therefore, the third aim of the current study was to test whether effects of distal prosody could be elicited with more varied intonation patterns. A case of particular interest is a downtrend pattern, characterized as a declining fundamental frequency (f0) with a low pitch at the end of a phrase or utterance. This low pitch is generally thought to signal finality (Lehiste, 1975; Streeter, 1978), or the end of a turn in discourse (Geluykens & Swerts, 1994; Schaffer, 1984; Swerts & Geluykens, 1994). Traditionally referred to as “declination,” where downtrend is used as a theory-neutral descriptive term of the phenomenon, the presence of declining f0 across a phrase has long been established for English, and is widespread across languages, including prosodically similar languages like Dutch, and languages with distinct intonation structures, such as Japanese (Cohen & 't Hart, 1968; Kubozono, 1989; Pike, 1945; van den Berg et al., 1992). In the Autosegmental-Metrical theory of intonation, the realization of the pitch contour within a phrase, including the falling f0 of declination, is the result of interpolation between “Low” and “High” tonal targets for phrasal tones and accents (Ladd, 2008; Liberman & Pierrehumbert, 1984; Pierrehumbert & Beckman, 1988; van den Berg et al., 1992). The perception of a pitch accent is not reliant on the presence of a certain fundamental frequency range alone (Pierrehumbert, 1979), and listeners must use other cues from the intonational context to categorize a pitch accent and gauge its location within a word. Therefore, the effects of distal prosody on the perception of word boundaries should be elicited even when the prosodic patterning is not signaled by repeated absolute frequencies.

2. Experiment 1

Dilley and McAuley (2008) have previously used compound word materials to show that prosodic patterning consisting of either distal f0 cues alone or distal duration cues alone affects perceived prosodic structure. However, it is not clear whether the morphological structure of compounds may lead to differences in lexical processing compared to other morphological structures; therefore, segmentation of ambiguously organized compound elements may not be reflective of general lexical processing. The aim of Experiment 1 was to examine, using a distinct set of stimuli with more typical morphological structure than used in Dilley and McAuley (2008), whether perceived prosodic structure could be altered by manipulating either f0 alone or duration alone, and to test whether combined f0 and duration cues yielded the strongest effects on perceived prosodic structure.

2.1. Methods

2.1.1. Participants and design

Sixty participants from Michigan State University completed the experiment in return for course credit. All participants were at least 18 years of age ($M=20.5$, $SD=3.4$), had self-reported normal hearing with varied number of years of formal music training ($M=3.3$,

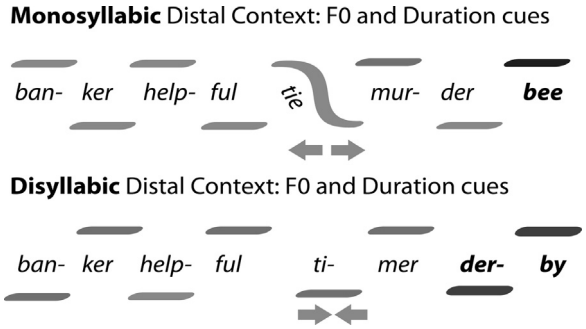

Fig. 1. Schematic of the Monosyllabic and Disyllabic Distal Contexts, with both f0 and Duration manipulations. Gray bars above the syllables represent the High (H) tones, while gray bars below the syllables represent the Low (L) tones. Outward pointing arrows under the fifth syllable in the Monosyllabic Distal Context indicate time expansion, and inward pointing arrows under the fifth syllable in the Disyllabic Distal Context indicate compression. Two tones were imposed on the fifth syllable in the Monosyllabic Distal Context to elicit the perception of the fifth syllable as its own tonal grouping. Note that across both Distal Contexts, the final three syllables bear the same intonation pattern.

SD=4.0). The design was a 2 (Distal Context: disyllabic, monosyllabic)×3 (Cue Type: f0, Duration, f0+Duration) mixed factorial. Cue Type was a between-subject factor, while Distal Context was a within-subject factor. Twenty participants were randomly assigned to each condition. One participant was not included in the final sample due to inattention to the task, resulting in $n=19$ for the f0 condition, $n=20$ for the Duration condition, and $n=20$ for the f0+Duration condition.

2.1.2. Materials

There were 30 experimental items and 90 filler items. Experimental items consisted of 30 syllable sequences of eight syllables each, corresponding to experimental items used in (Dilley et al., 2010) Experiment 1. Each sequence contained two disyllabic words, followed by a string of four syllables which could be parsed into words in two different ways (see Fig. 1 for a schematic of the stimuli, Fig. 2 for a representation of the waveform and pitch track, and Appendix for the complete set of stimuli). The example sequence in Figs. 1 and 2, /tɑɪmɜ˞dɜ˞bi/, can be perceived as containing either a monosyllabic final word, as in *tie murder bee*, *or a* disyllabic final word, *timer derby.* The syllable sequences consisted of lists of lexical items which were intended to be unrelated semantically. Note that each item occurred in both disyllabic and monosyllabic distal context conditions within an experimental condition, counterbalancing the pairing of disyllabic and monosyllabic distal context conditions with items across participants (see below). Since differences in frequency among the possible words formed by adjacent syllables in sequence were fixed within each item (i.e., the syllables in an item always occurred in the same order), and each item occurred in both distal context conditions, differences in frequency among possible words cannot be responsible for any observed effects of distal prosodic context on perception of prosodic structure. Filler items were made up of unambiguous lexical sequences which ranged in length from six to ten syllables and contained a mixture of monosyllabic and disyllabic words. Half of the fillers ended with monosyllabic words and half ended with disyllabic words.

All stimuli for the present experiment were generated from those used in (Dilley et al., 2010) Experiment 1A, which had included combined duration and f0 manipulations for disyllabic and monosyllabic Distal Contexts for each stimulus item. Manipulations used to generate Duration condition stimuli (which lacked f0 variation) and f0 condition stimuli (which lacked duration variation) are described below. Recordings for the original experiment of Dilley et al. (2010) had been made by a female, native speaker of American English and trained phonetician (author L.D.) who produced the syllable sequences with a monotone intonation and with as "neutral" articulation as possible, so as not to indicate any specific strengthened or stressed syllables.[1] However, the entire sequence was spoken as a fluent string of disyllabic words (and not as isolated syllables); therefore, the final syllable may be more likely to be interpreted as part of a disyllabic word than as a monosyllabic word. This possibility is addressed in the Discussion. Duration and f0 manipulations were conducted following methods of (Dilley & McAuley, 2008) and (Dilley et al., 2010). These procedures are described below.

2.1.2.1. Duration condition. For the Duration condition, filler items and experimental items in the disyllabic and monosyllabic Distal Context conditions were given a monotone f0 (205 Hz) using the overlap-add function in Praat (Boersma & Weenink, 2002). Thus, the disyllabic and monosyllabic Distal Context conditions differed only in durational properties. For the monosyllabic Distal Context, the fifth syllable region (i.e., the portion of each item ranging from the vowel onset of the fifth syllable to that of the sixth syllable) was lengthened by a factor of 1.8 also using the overlap-add function in Praat (Boersma & Weenink, 2002); this lengthening was expected to induce the perception of an additional "beat" on the fifth syllable and continue the perceptual groupings of syllables in the distal context (such that the lengthened fifth syllables would comprise its own "group") to result in a monosyllabic parse of the final syllable in the sequence (see Fig. 2a). A short (<25 ms) linear ramp was used to interpolate between surrounding temporally unmodified speech (which had a temporal modification factor of 1.0, leading to the same duration as the original). The expansion factor of 1.8 for the fifth syllable was chosen so that, in addition to being perceived as containing two beats, the fifth syllable duration and the durations of the surrounding expected groupings of syllables would be maximally isochronous, while maintaining naturalness and comprehensibility of the fifth syllable as determined by trained phoneticians (authors T. M. and L. D.). In contrast, because the original

[1] In addition, since the original materials from all experiments reported here were spoken with a monotone f0, there were no consistencies regarding f0 peak or valley alignment across the unaltered syllable sequences. Thus the utterances prior to f0 manipulation would be well-described as consisting of a single f0 "plateau" (Knight, 2003).

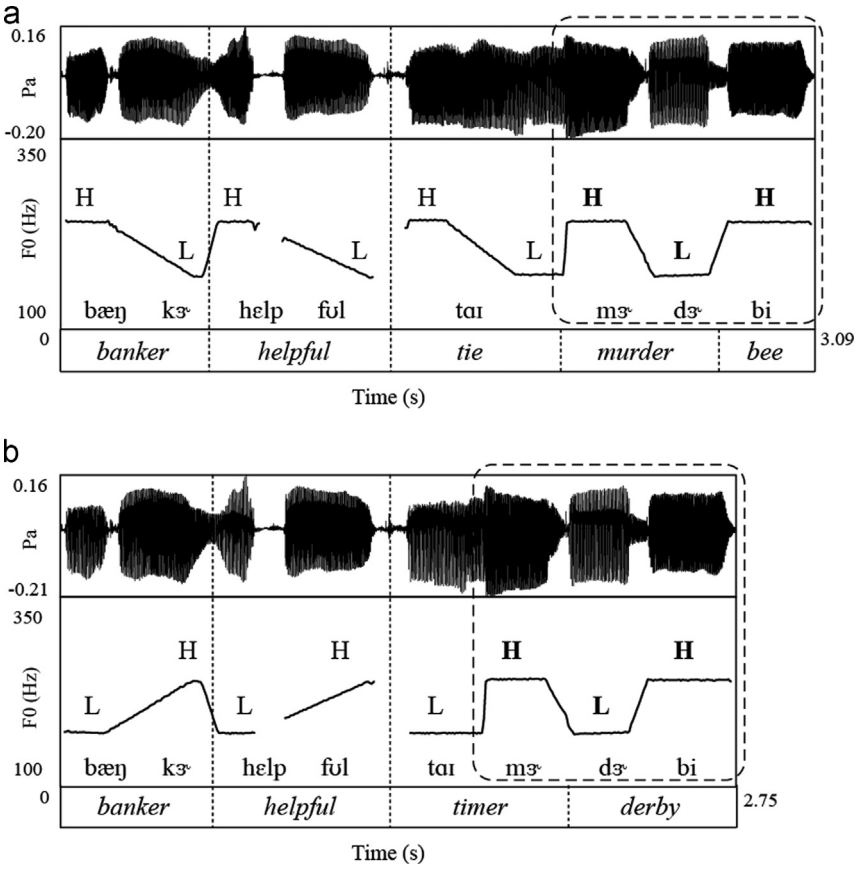

Fig. 2. Wave form and pitch track for an example stimulus item in the f0+Duration condition in Experiment 1 (LH Final Intonation pattern). Panel (a) shows an item in the monosyllabic distal context and panel (b) shows an item from the disyllabic distal context. The fifth syllable, *tie*, is lengthened and bears two tones in the monosyllabic distal context, but the acoustics of the final three syllables are identical for both (a) and (b).

production of the fifth syllable had been produced as slightly longer than the surrounding syllables to facilitate lengthening, for the disyllabic Distal Context, the fifth syllable region was shortened by a factor of 0.9; this was expected to induce perception of grouping of syllables into sequences of pairs (such that the fifth syllable was the first syllable of a disyllabic word) and to result in a disyllabic parse of the final syllable in the sequence (see Fig. 2b). The shortening of the fifth syllable was conducted so that the durations of the expected groupings of syllables would be maximally isochronous perceptually and elicit the intended grouping patterns. Again, a <25 ms linear ramp was used to interpolate between surrounding temporally unmodified speech and the temporally-modified fifth syllable region (which had a temporal modification factor of 1.8 or 0.9 in the monosyllabic or disyllabic Distal Context conditions, respectively).

2.1.2.2. f0 condition. The f0 manipulation consisted of intonation patterns imposed over the eight-syllable sequence, resulting in a series of low and high tones, with a three-syllable ending sequence of high–low–high (HLH) in both the monosyllabic and disyllabic Distal Contexts. The acoustic characteristics of the final three-syllable sequence were held acoustically constant across the disyllabic and monosyllabic Distal Context conditions. Each syllable of the stimulus items in the f0 condition was assigned a single H or L tone (or HL fall). For the disyllabic Distal Context, the initial five syllables had a LHLHL pattern with one tone per syllable. The f0 pattern in the disyllabic condition was expected to give rise to a (LH)(LH)(LH)(LH) grouping pattern across the eight-syllable sequence, with the final two syllables being grouped into an (LH) syllable sequence. For the monosyllabic Distal Context, the syllable sequence would start with (HL) pairs, and both a high and low tone would occur on the fifth syllable, resulting in an expected (HL)H grouping for the final three syllables; in this case, the final two syllables were not grouped together and the final H syllable stood alone (see Fig. 2). The fundamental frequency (f0) of the low tones was 165–175 Hz, and the high tones were at 235–245 Hz. f0 contours were stylized as a series of straight lines in Praat, thereby removing microprosodic variation. The target f0 was achieved by the rime of the syllable to which the tone was assigned. The f0 values of tones were selected so that H and L would sound distinct from one another while giving rise to an utterance rhythm that the first two authors judged to be unambiguous perceptually; moreover, we avoided selecting f0 values for H and L that formed a consonant melodic pitch interval (e.g., an interval of a third or a fourth).

In the f0 condition, durations of experimental items were adjusted in order to neutralize timing differences across the two Distal Context conditions of the original Dilley et al. (2010) stimuli. Neutralization of duration differences was accomplished by using the overlap-add function in the graphical user interface in Praat to shorten or lengthen the duration of the fifth syllable region of monosyllabic and disyllabic experimental items, respectively, from Dilley et al. (2010) to a value of 1.6 across all items (i.e., 1.6 times the duration of that portion of the original speech sound recording from which both the monosyllabic and disyllabic versions of each

item had been derived in the Dilley et al. 2010 materials).[2] The factor of 1.6 was selected since it was judged to result in a fifth syllable duration which would provide ambiguous temporal cues supporting either a monosyllabic or disyllabic rhythmic interpretation.

2.1.2.3. f0+Duration condition. Stimuli for the f0+Duration condition were the same as those used in Experiment 1A of Dilley et al. (2010). Syllable sequences were assigned a tonal structure as described above for the f0 condition. In both the monosyllabic and disyllabic Distal Contexts, the final three syllables were identical, with the final two syllables carrying a (LH) pitch pattern. This intonation contour was created by resynthesizing the third section of the sound file with an f0 of 165–175 Hz for the low (L) tones, and 235–245 Hz for the high (H) tones. For the duration manipulation in the f0+Duration condition, the fifth syllable in the monosyllabic condition was lengthened by a factor of 1.8, as described above, and in the disyllabic condition, the fifth syllable was compressed by a factor of 0.9. As described for the Duration condition, these manipulations were performed to elicit the perception of differing patterns of strong and weak syllables and syllable groupings across the monosyllabic and disyllabic Distal Contexts.

Filler sentences were created to be similar to experimental items in both the monosyllabic and disyllabic Distal Context conditions, with approximately half of the filler sequences consisting of a rising (LH) intonation pattern on each word, and the other half consisting of a falling (HL) pattern on each word. f0 values for high and low tones were in the range of 220–260 Hz and 150–190 Hz, respectively. The amplitude of all experimental and filler sequences was normalized to 70 dB SPL.

2.1.3. Equipment

Stimulus presentation and response collection were controlled by Eprime 1.1 (Psychology Software Tools, Inc., Pittsburgh, PA) running on a Dell Optiplex GX620 desktop computer. Participants listened to stimuli over Sennheiser HD 280 Pro headphones and typed their responses using the computer keyboard.

2.1.4. Procedure

On each trial, participants listened to a syllable sequence and reported the last word they heard in the sequence. Participants initially completed 16 practice trials, which consisted of only filler sequences. Participants then completed test trials, which consisted of 30 experimental sequences and 90 filler sequences. Sequences on the test trials were pseudo-randomly ordered so that no two experimental items occurred on successive trials. Half of the 30 experimental sequences were given a disyllabic Distal Context, and the other half were given a monosyllabic Distal Context, as described above. Pairing of specific syllable sequences with Distal Context was counterbalanced across sequences to create two lists. Two additional lists were then created by reversing the order of presentation of the sequences, for a total of four unique lists. Equal numbers of participants were randomly assigned to each list. Participants' typed responses were coded as monosyllabic or disyllabic; uninterpretable responses, those in which spelling or typing errors resulted in responses that did not correspond to English lexical items, were removed from analysis (14 items, or less than 1% of the data, were removed).

2.2. Results

Fig. 5 shows mean proportions of disyllabic responses for the experimental sequences in disyllabic and monosyllabic Distal Contexts for the Duration, f0, and f0+Duration conditions. A mixed effect logistic regression model (e.g., Jaeger, 2008) was fitted to the data in the R statistical programming language (Bates, Maechler, & Bolker, 2012), with Distal Context and Cue Type as fixed effects, and Subject and Item as random effects (Table 1). A full random effects structure was used, with random slopes for the within Subject and within Item manipulations (Distal Context for Subjects and Distal Context and Cue Type for Items, as well as the interaction term between Distal Context and Cue Type for Items). Distal Context was a significant predictor of disyllabic responses, with fewer disyllabic responses in the monosyllabic context than in the disyllabic context ($\beta = -2.3706$, $SE = 0.3687$, $z = -6.429$, $p < 0.001$). The Cue Type of f0+Duration leads to significantly higher rates of disyllabic responses ($\beta = 2.7054$, $SE = 0.6176$, $z = 4.381$, $p < 0.001$). The lowest rates of disyllabic responses occurred in the monosyllabic Distal Context for the f0+Duration Cue Type, and the interaction was significant ($\beta = -3.2690$, $SE = 0.6746$, $z = -4.846$, $p < 0.001$). The best model fit was obtained with both the Distal Context and Cue Type predictors, with significantly worse fit without Distal Context ($\chi = 529.73$, $p < 0.001$) or without Cue Type ($\chi = 38.019$, $p < 0.05$). Akaike Information Criterion scores (Akaike, 1973) also showed the best fit for the model with both Distal Context and Cue Type as predictors (AIC = 1622.7)

Next, signal detection measures d' and c were used to distinguish between participants' sensitivity to the distal prosody manipulation from any general tendency to make a disyllabic or monosyllabic response (see MacMillan & Creelman, 1991). Signal detection measures can be used to consider participants' sensitivity to the Distal Context manipulation separately from possible response biases to report either monosyllabic or disyllabic final words. For the signal detection analysis, the report of a disyllabic final word in the disyllabic Distal Context was coded as a hit, while the report of a disyllabic final word in the monosyllabic Distal Context

[2] The duration manipulations applied in the f0 Cue Type condition were intended to "undo" the duration manipulations that had been imposed in the Dilley et al. (2010) materials. Specifically, the monosyllabic Distal Context condition items of the f0 Cue Type condition were created by shortening the fifth syllable region from the monosyllabic condition of the Dilley et al. (2010) materials by 0.89; this syllable had been lengthened by Dilley et al. (2010) by a factor of 1.8 (1.6/1.8≈0.89). Moreover, the disyllabic Distal Context condition items of the f0 Cue Type condition were created by lengthening the fifth syllable region from the disyllabic condition of the Dilley et al. (2010) materials by 1.8; this syllable had been shortened by Dilley et al. (2010) by a factor of 0.9 (0.9×1.8≈1.6). Durational differences between monosyllabic and disyllabic Distal Context conditions remaining after these manipulations were negligible; on average, monosyllabic Distal Context condition files were 2 ms longer than disyllabic Distal Context condition files, a value which was not significantly different from zero in a single-sample *t*-test, $t(29) = 1.014$, $p = 0.319$.

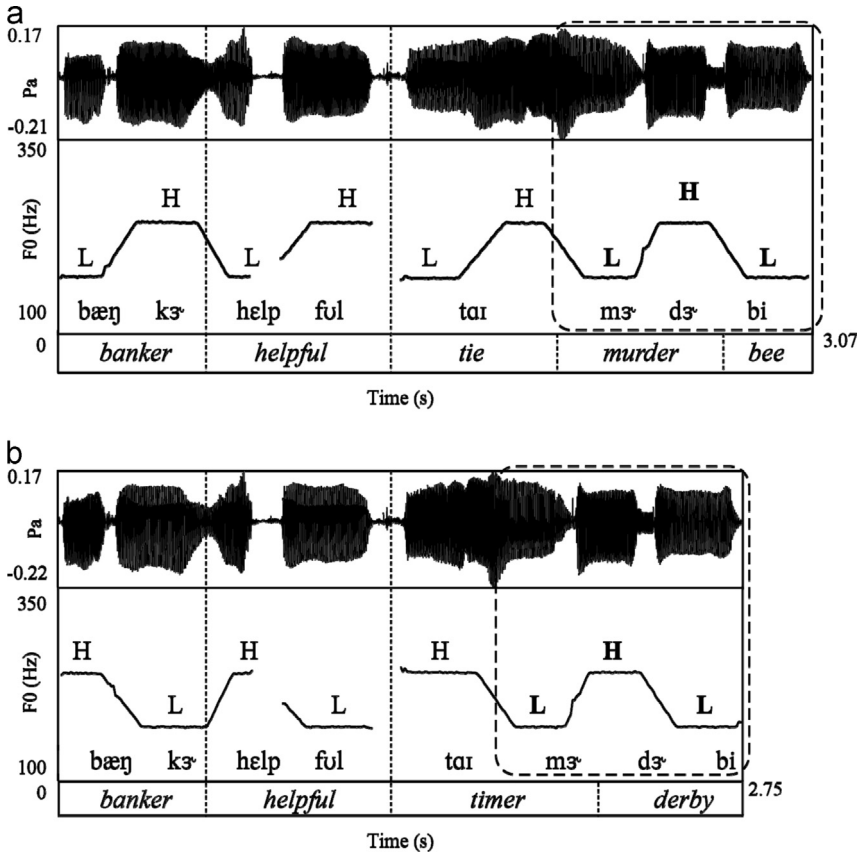

Fig. 3. Wave form and pitch track for an example stimulus item in the f0+Duration condition in Experiment 2 (HL Final Intonation pattern). Panel (a) shows an item in the monosyllabic distal context and panel (b) shows an item from the disyllabic distal context. The fifth syllable, *tie*, is lengthened and bears two tones in the monosyllabic distal context, but the acoustics of the final three syllables are identical for both (a) and (b).

was coded as a false alarm. Hits and false alarms for each participant and for each item were used to calculate the signal detection measures d' (a measure of sensitivity to Distal Context calculated as $d' = z(\text{Hit rate}) - z(\text{False alarm rate})$) and c (a measure of response bias calculated as $c = -0.5(z(\text{Hit rate}) + z(\text{False alarm rate}))$). d' scores provided a standardized measure of the degree to which final word reports were both (1) affected by the Distal Context and (2) showed perceptual organization into precisely the predicted groupings. $d' = 0$ would then indicate no effect of Distal Context on final word reports; likewise, the more positive the value of d', the greater the extent to which the predicted groupings were observed (as opposed to the opposite pattern of groupings). Values of c provided a standardized measure of any bias in final syllable groupings, with values greater than or less than zero indicating an overall tendency to respond with monosyllabic or disyllabic final words, respectively, regardless of type of Distal Context.

Table 2 (first row) shows mean values and standard deviations of d' and c for the Duration, f0, f0+Duration conditions. In the f0+Duration condition, values of d' (sensitivity Distal Context) were significantly higher ($M = 1.83$, $SD = 0.76$) than in the Duration ($M = 0.99$, $SD = 0.69$) or f0 ($M = 1.05$, $SD = 0.63$) conditions. An ANOVA on d' with Cue Type (Duration, f0 and f0+Duration) as the independent variable showed an effect of Cue Type (F_1 (2,56) $= 9.06$, $p < 0.001$, $\eta^2 = 0.245$; F_2 (2,87) $= 10.61$, $p < 0.001$, $\eta^2 = 0.196$; $MinF'(2, 131) = 4.89$, $p < 0.01$). With respect to the response criterion, c, there was a general tendency to respond with the disyllabic word ($M = -0.41$, $SD = 0.38$); the criterion was significantly different from zero, t (1,59) $= -8.38$, $p < 0.001$. An ANOVA on c revealed no main effect of Cue Type (F_1 (2,56) $= 0.85$, $p = 0.435$, $\eta^2 = 0.029$; F_2 (2,87) $= 0.29$, $p = 0.75$, $\eta^2 = 0.007$; $MinF'(1, 132) = 0.21$, $p = 0.81$).

2.3. Discussion

Results of Experiment 1 replicated the effect of distal prosodic patterning on perceived prosodic structure first reported in Dilley and McAuley (2008), extending these results to a more typical morphological structures. In particular, these results demonstrate that perceived prosodic structure can be altered by either f0 cues alone or duration cues alone. Across all three cue type conditions, Distal Context was a significant predictor of participants' reporting of disyllabic final words, with higher report rates in the disyllabic Distal Context condition than in the monosyllabic Distal Context, even though the final syllables were acoustically identical across Distal Context conditions for each item.

Sensitivity to the difference between the disyllabic and monosyllabic Distal Contexts as indicated by d' scores tended to be slightly lower in the conditions in which duration and pitch were manipulated independently. This pattern of results is compatible with those of (Dilley et al., 2010), showing that a combination of f0 and duration information in the distal context comprise a strong cue for the perceptual organization of ambiguous lexical sequences. However, the current results confirm that both f0 and duration information function as effective cues to prosodic structure. Thus, the presence of either cue can affect the perception of prosodic structure and

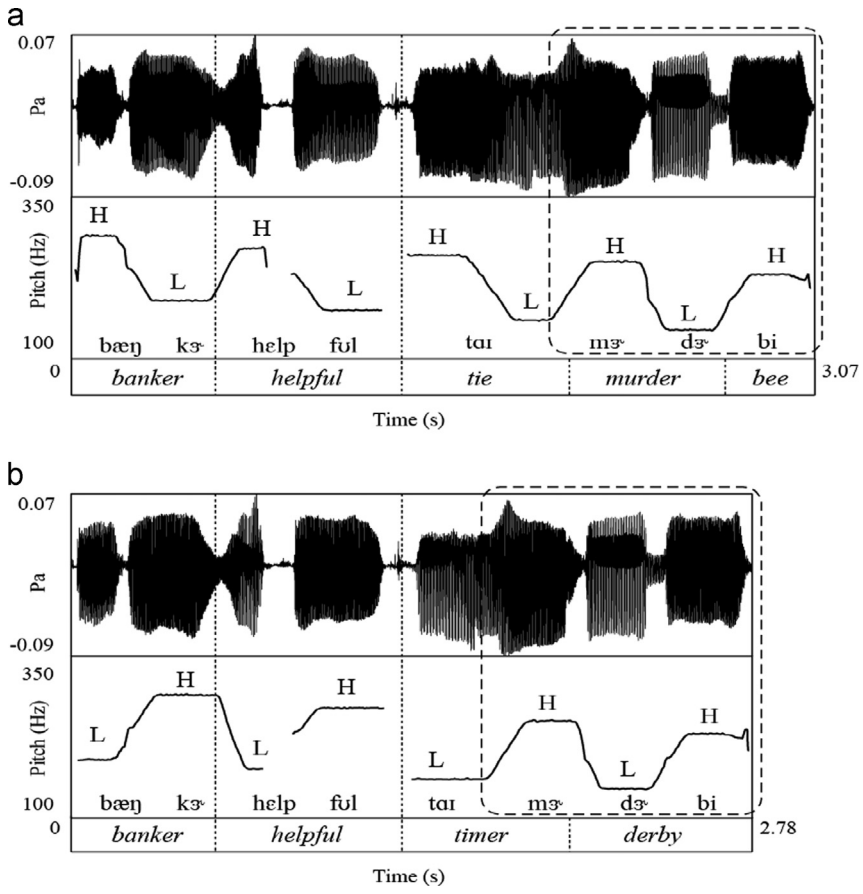

Fig. 4. Wave form and pitch track for an example stimulus item in Experiment 3 ("downtrend" intonation pattern). Panel (a) shows an item in the monosyllabic distal context and panel (b) shows an item from the disyllabic distal context. The fifth syllable, *tie*, is lengthened and bears two tones in the monosyllabic distal context, but the acoustics of the final three syllables are identical for both (a) and (b).

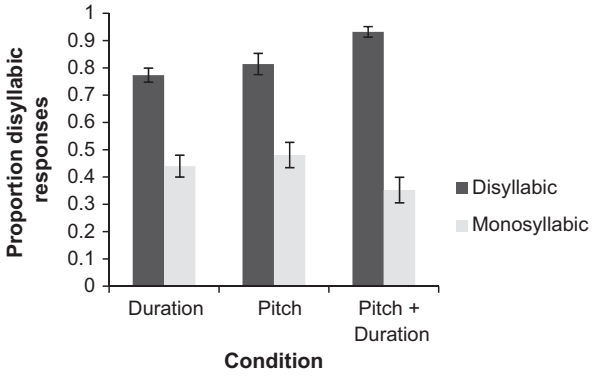

Fig. 5. Mean proportions of disyllabic responses in Experiment 1 for the LH Final pattern as a function of cue type in disyllabic and monosyllabic distal contexts.

Table 1
Logit mixed effects model for predictors of disyllabic words reports in Experiment 1.

	Estimate	Std. error	*z* Value	*p* Value
(Intercept – *disyllabic, Duration Cue Type*)	1.9978	0.4588	4.3549	***p*<0.001**
Context – *monosyllabic*	−2.3706	0.3687	−6.429	***p*<0.001**
Cue Type – *f*0	0.4804	0.4431	1.084	$p=0.278$
Cue Type – *f*0+*Duration*	2.7054	0.6176	4.381	***p*<0.001**
Context – *monosyllabic* * Cue Type – *f*0	−0.5026	0.4684	−1.073	$p=0.283$
Context – *monosyllabic* * Cue Type – *f*0+*Duration*	−3.2690	0.6746	−4.846	***p*<0.001**

Table 2
Means and standard deviations for d' and c scores for Experiments 1 and 2 for the f0, Duration, and f0 + Duration conditions.

	Duration		f0		f0 + Duration	
	d'	c	d'	c	d'	c
Experiment 1: LH Final Intonation	0.99 (*0.69*)	−0.32 (*0.31*)	1.05 (*0.63*)	−0.44 (*0.48*)	1.83 (*0.76*)	−0.47 (*0.34*)
Experiment 2: HL Final Intonation	0.83 (*0.90*)	−0.82 (*0.34*)	1.04 (*0.80*)	−0.82 (*0.35*)	1.33 (*0.86*)	−0.81 (*0.40*)

the grouping of syllables into words. This result has important implications for our understanding of speech perception processes, as the presence of prosodic cues in the acoustic speech stream is highly variable – the ability of listeners to group syllables into words on the basis of either f0 and duration cues independently indicates flexibility in the perceptual system for determining the structure of the linguistic material in the speech stream. In this experiment, when participants gave final disyllabic word reports, they perceived a relatively large prosodic boundary (Prosodic Word or higher-level) in proximal material just prior to the final two syllables, as opposed to between them. In the case of disyllabic word reports, the tonal and temporal patterning in the distal prosodic context led listeners to hear the final two syllables as grouped together and as a single prosodic unit. In contrast, when participants gave final monosyllabic word reports, they perceived a relatively large prosodic boundary in proximal material just prior to the final syllable. In both cases, the implied prosodic structuring of proximal material was parallel to that of the distal context, as indicated by either f0, duration, or combined f0 and duration prosodic cues. The presence of a bias towards disyllabic word reports as indicated by the measure *c* is not unexpected given the characteristics of the originally uttered sequences, which consisted of a fluently spoken series of disyllabic words. Therefore, in the monosyllabic Distal Context condition, the perception of a prosodic boundary preceding the final syllable, eliciting a monosyllabic final word report, represents an altered perception affected by the distal prosodic context.

The results of Experiment 1 are consistent with the hypothesis that prosodic patterns present in the distal context induce the grouping of syllables in a manner consistent with general principles of auditory perceptual organization. The effects of f0 and duration cues in combination yielded reliably larger effects than each cue in isolation; however, either f0 cues alone or duration cues alone present in the signal can influence the grouping of syllables into higher-level prosodic structures. These results confirm that distal prosodic patterning in a single acoustic dimension can affect perception of proximal prosodic structures for speech with more typical morphological structure than that used in Dilley and McAuley (2008).

3. Experiment 2

In Experiment 1 in the f0 and f0 + duration Cue Type conditions, a LHLHLHLH tonal pattern was used for the disyllabic Distal Context items (corresponding to predicted LH perceptual groupings), whereas a HLHLHLHLH tonal pattern was used for the monosyllabic Distal Context items (corresponding to predicted perceptual groupings of HL). As an initial investigation of whether distal prosodic effects generalize to other intonational environments, a second experiment was conducted where the opposite tonal pattern (i.e., repeated HL sequences in the disyllabic Distal Context) was assigned (Fig. 3). If effects of distal prosody generalize to the opposite tonal pattern, then similar to Experiment 1, the disyllabic Distal Context should elicit disyllabic final word reports, and monosyllabic Distal Contexts should elicit monosyllabic final word reports.

3.1. Methods

3.1.1. Participants and design

Seventy-one native speakers of American English ($M = 21.7$ years, $SD = 6.3$ years) from the Michigan State University community completed the experiment in return for course credit or nominal financial compensation. Participants self-reported normal hearing and varied in number of years of formal music training ($M = 2.9$ years, $SD = 3.6$ years). The design of the experiment was a 2 (Distal Context: disyllabic, monosyllabic) × 3 (Cue Type: Duration, f0, f0 + Duration) mixed-factorial. Cue Type was varied between subjects, while Distal Context was varied within subjects. Participants were randomly assigned to one of the three Cue Type conditions. Four participants were not included in the final data set due to inattention to the task. This resulted in the following numbers of participants in each level of Cue Type: Duration ($n = 21$), f0 ($n = 22$), and f0 + Duration ($n = 24$).

3.1.2. Materials

Experiment 2 used the same 30 experimental items and 90 filler items as in Experiment 1. Stimuli in the duration and pitch manipulation conditions were then constructed following procedures similar to those used in Experiment 1. The distal prosodic contexts were created from the same set of original recordings as had been used to generate stimuli in Experiment 1.

3.1.2.1. Duration condition. For the Duration condition, stimuli were the same as in Experiment 1, where each sound file had been resynthesized to a monotone f0.

3.1.2.2. f0 condition. Stimuli in the f0 condition were created following the procedures used for pitch manipulations in Experiment 1. The f0 of the low tones was 165–175 Hz, and the high tones were at 235–245 Hz, consisting of the same fundamental frequencies as

those of the contour used in Experiment 1, but with low (L) and high (H) tones occurring in the opposite order. Thus, the final two syllables consisted of a HL pattern, whereas in Experiment 1, the final two syllables consisted of a LH pattern; as in Experiment 1, the final syllables were acoustically identical across the disyllabic and monosyllabic Distal Context conditions. In the monosyllabic Distal Context of the f0 condition in Experiment 2, the fifth syllable consisted of a low tone transitioning to a high tone (L–H).

As in Experiment 1, the fifth syllable was lengthened by a factor of 1.6, giving rise to a fifth syllable duration which would provide ambiguous temporal cues supporting either a monosyllabic or disyllabic rhythmic interpretation (see Fig. 3). The duration manipulation was performed on both the monosyllabic and disyllabic Distal Contexts, so that the duration of the fifth syllable was matched across both conditions, and the only prosodic cue which varied was f0.

3.1.2.3. f0+Duration condition. The stimuli in the f0+Duration condition contained both the f0 manipulation performed as described above, and the duration manipulation as described for Experiment 1. Thus, the fifth syllable in the monosyllabic context consisted of the (L–H) pitch pattern, and was lengthened by a factor of 1.8 from its original duration. In the disyllabic context, the fifth syllable contained only a high pitch (H), and the duration of the syllable was altered by a factor of 0.9. As in the other Cue Type conditions, the final three syllables were acoustically identical across the monosyllabic and disyllabic Distal Context. The portion selected as the fifth syllable was identical to the portion selected in both the Duration and f0 conditions, and the manipulations were performed by resynthesizing the file with both duration and f0 manipulations concurrently, using the overlap-add function in Praat. Filler sentences were the same as those used in Experiment 1, with approximately half of the filler sequences consisting of a rising (LH) intonation pattern on each word, and the other half consisting of a falling (HL) pattern on each word. The amplitude of all experimental and filler sequences was normalized to 70 dB SPL.

3.1.3. Procedure

All aspects of the procedure and equipment were identical to Experiment 1. Any uninterpretable responses were removed from analysis (4 items, or less than 0.01% of the data, were removed).

3.2. Results

Fig. 6 shows mean proportions of disyllabic responses for the disyllabic and monosyllabic Distal Contexts for the Duration, f0, and f0+Duration conditions. A mixed effect logistic regression model (Table 3) with Distal Context and Cue Type as fixed effects, and Subject and Item with a full random effects structure (random slopes for within Subject and within Item manipulations, as in Experiment 1) revealed that Distal Context was a significant predictor of disyllabic responses, with fewer disyllabic responses in the monosyllabic context than in the disyllabic context ($\beta = -1.9107$, $SE = 0.3844$, $z = -4.970$, $p < 0.001$). The Cue Type of f0+Duration leads to significantly higher rates of disyllabic responses than in the Duration condition ($\beta = 1.4456$, $SE = 0.4479$, $z = 3.227$, $p < 0.01$), while the difference between Cue Type conditions of f0 and of Duration was not significant, ($\beta = 0.6282$, $SE = 0.3911$, $z = 1.606$, $p = 0.1083$). The lowest rates of disyllabic responses occurred in the monosyllabic Distal Context for the f0+Duration Cue Type, and the interaction between Distal Context and the f0+Duration Cue Type was significant, ($\beta = -2.0853$, $SE = 0.5864$, $z = -3.556$, $p < 0.001$), confirming that the effect of Distal Context was greatest in this condition. Model fit was significantly worse without Distal Context as a predictor ($\chi = 416.74$, $p < 0.001$). Fit was slightly better without Cue Type than with both Cue Type and Distal Context ($\chi = 35.913$, $p < 0.05$); however, Akaike Information Criterion scores (Akaike, 1973) showed the fit for the model with both Distal Context and Cue Type as predictors (AIC = 1608.7), was only slightly higher than for the model with only Distal Context as a predictor (AIC = 1600.6). The full model with both fixed effects is reported here (Table 3).

Table 2 shows mean values and standard deviations of d' and c for the Duration, f0, f0+Duration conditions. Consistent with Experiment 1, values of d' were highest for the f0+Duration condition ($M = 1.33$, $SD = 0.86$), next highest for the f0 condition ($M = 1.04$, $SD = 0.80$), and lowest for Duration condition ($M = 0.83$, $SD = 0.90$); an effect of Cue Type was significant only for items ($F_1(2,64) = 2.01$, $p = 0.14$, $\eta^2 = 0.06$; F_2 (2, 87) = 6.85, $p < 0.005$, $\eta^2 = 0.14$; $minF'(2,101) = 1.56$, $p = 0.22$). Post-hoc pair-wise

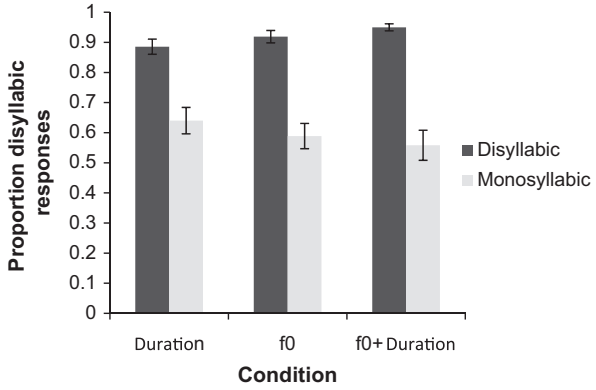

Fig. 6. Mean proportions of disyllabic responses in Experiment 2 for the HL Final pattern as a function of cue type in disyllabic and monosyllabic distal contexts.

Table 3
Logit mixed effects model for predictors of disyllabic words reports in Experiment 2.

	Estimate	Std. error	*z* Value	*p* Value
(Intercept – *disyllabic, Duration Cue Type*)	2.8610	0.3823	7.483	***p*<0.001**
Context – *monosyllabic*	−1.9107	0.3844	−4.970	***p*<0.001**
Cue Type – *f*0	0.6282	0.3911	1.606	$p=0.1083$
Cue Type – *f*0+*Duration*	1.4456	0.4479	3.227	***p*<0.01**
Context – *monosyllabic* *Cue Type – *f*0	−0.9541	0.5119	−1.864	$p=0.0623$
Context – *monosyllabic* *Cue Type – *f*0+*Duration*	−2.0853	0.5864	−3.556	***p*<0.001**

comparisons for items revealed that the difference between the f0+Duration condition and the Duration condition was significant, $p<0.005$. With respect to the response criterion, *c*, there was a general tendency to produce more disyllabic responses than monosyllabic responses ($M=-0.82$, $SD=0.36$); the criterion was significantly different from zero, $t(1,66)=-18.82$, $p<0.001$. An ANOVA on *c* revealed no significant effect of Cue Type, ($F_1(2,64)=0.01$, $p=0.99$, $\eta^2<0.001$; $F_2(2,87)=0.31$, $p=0.737$ $\eta^2=0.01$; minF' $(2,67)=0.01$, $p=0.99$).

3.3. Discussion

Results of Experiment 2 extend previous findings by showing that effects of distal prosody on the perception of prosodic structures generalize to a distinct intonation pattern which was not considered in previous work, but which is also a common pattern in English and other languages (Dilley et al., 2010; Dilley & McAuley, 2008). Both the intonation pattern of Experiment 1 (ending in a LH tonal sequence) and the intonation pattern of Experiment 2 (ending in a HL tonal sequence) produced final word reports consistent with predictions for the distal prosodic patterning and its effects on perceived prosodic structure. Although sensitivity to the prosodic manipulation appears to have been slightly lower in Experiment 2 than in Experiment 1 (based on *d'* scores), the results of Experiment 2, with Distal Context as a significant predictor of disyllabic word reports demonstrate that effects of distal prosodic patterning are not dependent on proximal material exhibiting a single, specific tonal contour type. When either duration or f0 cues in the distal context were manipulated to elicit a disyllabic word response to a lexically ambiguous sequence, participants' perception was affected in the predicted direction, with disyllabic Distal Contexts eliciting more disyllabic word reports than monosyllabic Distal Contexts. As in Experiment 1, for this to have occurred, participants had to have heard a relatively large prosodic boundary preceding the final two syllables, rather than as separating them. The effect of Distal Context was also consistent across Cue Type in Experiment 2, as in Experiment 1. In both experiments, there was a tendency for the effect of distal prosody to be strongest when f0 and duration cues were combined, but in Experiment 2, this effect was reliable for only the item analyses; crucially, f0 alone or duration alone were both effective cues to prosodic patterning and perceived prosodic structure.

4. Experiment 3

Experiments 1 and 2 showed that with intonation contours that varied in the sequencing of H and L tones, distal prosodic patterning influenced the perception of ambiguous lexical structures in later occurring acoustically identical contexts. However, in natural speech, phonologically defined high and low tones often show a characteristic lowering of fundamental frequency over the course of an utterance. Experiment 3 extends this investigation of the effects of distal prosody to more naturalistic intonation contours that exhibit a pattern of f0 downtrend, to test whether the effects can be elicited without the repetition of identical or nearly-identical absolute f0 changes.

4.1. Methods

4.1.1. Participants and design

Forty-one adult, native speakers of American English ($M=23.6$ years, $SD=6.4$ years) with self-reported normal hearing from the Michigan State University community received nominal financial compensation for their participation. Participants had self-reported normal hearing and varied in years of formal music training ($M=2.2$ years, $SD=2.8$ years). The design of the experiment was a 2 (Final Intonation: LH vs. HL)×2 (Distal Context: disyllabic vs. monosyllabic) mixed factorial. Final Intonation was a between-subject factor, while Distal Context was a within-subjects factor. Twenty-one participants were randomly assigned to LH Final Intonation condition, and 20 participants to the HL Final Intonation condition. Combined f0 and Duration cues were used to instantiate prosodic patterning in distal speech context, as the combination of cues had yielded the largest prosodic context effects in Experiments 1 and 2 and presented a more naturalistic manipulation due to their covariance in signaling prosodic structure in spoken language (Shattuck-Hufnagel & Turk, 1996; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992).

4.1.2. Materials

Stimuli for Experiment 3 consisted of the same eight-syllable lexical sequences from Experiments 1 and 2, with 30 experimental items and 90 filler items. In order to mimic the phonetic realization of downtrend contours found in natural speech utterances, an intonation pattern of decreasing f0 values on each successive high and low tone was imposed over the eight syllables. This resulted in tonal groupings that are higher at the beginning of the utterance and lower at the end of the utterance (see Fig. 4). In addition, the overall pitch range decreased as the utterance progressed, so that the largest differences between high and low tones occurred at the beginning of the utterance, and the difference between high and low tones was smaller at the end of the utterance, as is common in natural speech (e.g., Cohen, Collier, & 't Hart, 1982). The pattern of decreasing f0 was created by basing the sequence of tones on the intonation contours of both Experiments 1 and 2, beginning the disyllabic Distal Contexts with LH and HL pitch patterns, respectively. Stimuli for Experiment 3 were created from the same original recordings as used in Experiments 1 and 2. The pairing of specific lexical sequences with levels of Distal Context (monosyllabic vs. disyllabic) was counterbalanced across sequences to create two lists. Equal numbers of participants were randomly assigned to each list.

4.1.2.1. LH Final Intonation condition. For each eight-syllable lexical sequence in the LH condition, the pitch values ranged from 290 Hz (on the first high tone syllable) down to 145 Hz (on the last low tone syllable). In the disyllabic Distal Context, the sequence began with a low tone of 190 Hz, and each successive low tone dropped 15 Hz, whereas each successive high tone dropped 20 Hz. This pattern is representative of a typical downtrend pattern (see Fig. 4). In the monosyllabic Distal Context, in which the sequence began with a high tone, the first high tone was at 290 Hz, and the first low tone was at 190 Hz. Across both monosyllabic and disyllabic Distal Contexts, the final three syllables were acoustically identical, both exhibiting the downtrend pattern. The f0 values on the final three syllables (a HLH sequence) were 250 Hz, 145 Hz, and 230 Hz, respectively. In addition, to mimic the final pitch fall that occurs on utterance-final sonorants in natural speech, an f0 drop of 15 Hz was imposed 100 ms from the end of the final syllable in the sequence, so that the ending f0 was at 215 Hz. This minor f0 drop was added to increase the perceived naturalness of the intonation contour.

The duration manipulation was performed in the same manner as in Experiment 2; the selection of the portion to be manipulated was also conducted in the manner described for Experiment 2. The duration of the fifth syllable was increased by a factor of 1.8 in the monosyllabic Distal Context condition, with 10 ms ramps at the beginning and end of the selected portion encompassing the duration manipulation. In the disyllabic condition, the fifth syllable was shortened by multiplying its duration by 0.9. The duration manipulations and f0 manipulations were performed concurrently using the overlap-add function in Praat.

4.1.2.2. HL Final Intonation condition. The pitch range of the HL condition was the same as that of the LH condition, with the highest high tone being 290 Hz, and the lowest low tone being 145 Hz (see Fig. 4). Across the intonation contour of both Distal Contexts in the HL condition, successive high tones dropped 20 Hz each and successive low tones dropped 15 Hz. As in the LH condition, the final three syllables in both the monosyllabic and disyllabic Distal Contexts were acoustically identical. The f0 values of these three syllables (an LHL sequence) were at 160 Hz, 230 Hz, and 145 Hz, respectively. The final f0 drop, imposed 100 ms before the end of the last syllable, was from 145 Hz to 130 Hz.

The duration manipulation in the HL-final condition was identical to that of the LH final condition. The duration of the fifth syllable was increased by a factor of 1.8 in the monosyllabic Distal Context condition, and shortened by a factor of 0.9 in the disyllabic condition, following methods in Experiment 2. The duration manipulations and f0 manipulations were performed concurrently using the overlap-add function in Praat to resynthesize the sound file.

Filler items were created by imposing a downtrend pattern over each syllable sequence such that the pitch range for the fillers was identical to that of the stimulus items (from 145 Hz to 290 Hz). Because the filler sequences ranged in length from six to ten syllables, the pitch fall for each successive high or low tone varied slightly across the filler items; this allowed the pitch range (from the lowest tone to the highest) across filler and experimental items to remain identical. Approximately half of the filler items began with a high tone on the first syllable of the sequence, and half began with a low tone on the first syllable. Fillers and stimuli were normalized to 58 dB SPL.

4.1.3. Procedure

The procedure in Experiment 3 was identical to that of Experiments 1 and 2. Participants' uninterpretable responses were removed from analysis (13 items, or approximately 1% of the data).

4.2. Results

Fig. 7 shows mean proportions of disyllabic responses for the disyllabic and monosyllabic Distal Contexts for the LH and HL Final Intonation conditions. A mixed effect logistic regression model (Table 4) with Distal Context and Final Intonation as fixed effects, and Subject and Item with a full random effects structure (random slopes for within Subject and within Item manipulations, as in Experiments 1 and 2) revealed that Distal Context was a significant predictor of disyllabic responses, with fewer disyllabic responses in the monosyllabic context than in the disyllabic context ($\beta = -5.1272$, $SE = 0.4932$, $z = -10.395$, $p < 0.001$). The HL Final Intonation condition led to slightly higher rates of disyllabic responses, but this difference was not significant ($\beta = 0.6897$, $SE = 0.6348$, $z = 1.086$, $p = 0.277$). The interaction between Final Intonation and Distal Context was not significant ($\beta = 0.6641$, $SE = 0.6956$, $z = 0.955$, $p = 0.340$). Model fit without Distal Context as a predictor was significantly worse than with it ($\chi = 532.69$, $p < 0.001$). Model fit with only

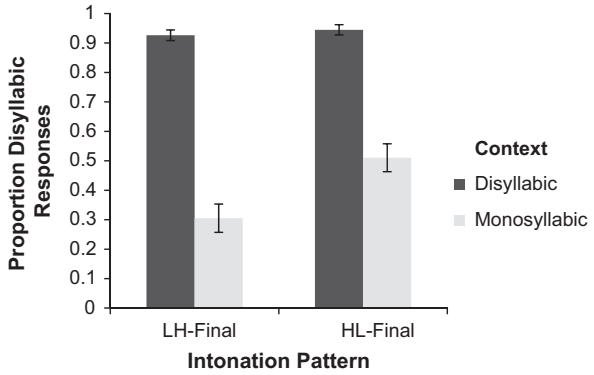

Fig. 7. Mean proportions of disyllabic responses in Experiment 3 for LH Final and HL Final downtrend intonation patterns in disyllabic and monosyllabic distal contexts.

Table 4
Logit mixed effects model for predictors of disyllabic words reports in Experiment 3.

	Estimate	Std. error	*z* Value	*p* Value
(Intercept – *disyllabic, HLH Intonation*)	3.8400	0.5027	7.639	**$p<0.001$**
Intonation – *LHL*	0.6897	0.6348	1.086	$p=0.277$
Context – *monosyllabic*	−5.1272	0.4932	−10.395	**$p<0.001$**
Context – *monosyllabic* *Intonation – *LHL*	0.6641	0.6956	0.955	$p=0.340$

Table 5
Mean value of *d′* and *c* with standard deviation in parentheses for Experiment 3.

LH Final Intonation		HL Final Intonation	
d′	***c***	***d′***	***c***
1.99 (*0.81*)	−0.39 (*0.36*)	1.45 (*0.76*)	−0.73 (*0.30*)

Distal context as a predictor was slightly better than with both Distal Context and Final Intonation, with an AIC score of 936.23 for only Distal Context and 936.31 with both predictors ($\chi=17.924$, $p<0.05$). The full model with both fixed effects is presented in Table 4.

Table 5 shows mean values of *d′* and *c* for the HL and LH Final Intonation conditions. A one-way ANOVA on *d′* revealed a significant effect of Final Intonation, ($F_1(1,40)=4.84$, $p<0.05$, $\eta^2=0.110$; $F_2(1,58)=9.01$, $p<0.01$, $\eta^2=0.14$; min$F'(1,79)=3.15$, $p=0.08$). Consistent with the analysis of response proportions, perceptual sensitivity was higher for the LH condition ($M=1.99$, $SD=0.81$) than for the HL condition ($M=1.45$, $SD=0.76$). With respect to the response criterion, *c*, participants gave fewer disyllabic responses with the LH Final Intonation pattern ($M=-0.39$, $SD=0.36$) than with the HL pattern ($M=-0.73$, $SD=0.30$) ($F_1(1,40)=10.72$, $p<0.005$, $\eta^2=0.22$; $F_2(1,58)=5.49$, $p<0.05$, $\eta^2=0.09$; min$F'(1, 96)=3.63$, $p=0.06$); in both conditions, the criterion was significantly different from zero ($p<0.001$).

4.3. Discussion

Results of Experiment 3 show that distal prosodic patterning in the context of a “downtrend” intonation pattern affected the perceived prosodic structure of proximal speech material in the predicted manner. In particular, disyllabic Distal Contexts, which were predicted to elicit groupings of the final two syllables into a single disyllabic word, yielded more disyllabic final word reports than monosyllabic Distal Contexts, which yielded a predominance of monosyllabic final word reports. These effects were obtained in both LH Final Intonation and HL Final Intonation patterns – disyllabic words reports were higher in the disyllabic condition regardless of whether the final two syllables exhibited a Low–High tonal pattern or a High–Low tonal pattern. These results demonstrate that the effect of distal prosody in eliciting perceived groupings of syllables is not dependent on a repeated pattern of identical f0 values. The main effect of Final Intonation Pattern (LH Final Intonation vs. HL Final Intonation) and the interaction between Final Intonation Pattern and Distal Context indicate that effects of distal prosody are somewhat stronger for the LH pattern than for the HL pattern. Similar to the first two experiments, there was a bias to report disyllabic final words, with a somewhat stronger bias for the HL pattern

than for the LH pattern, possibly reflecting differences in the inherent acoustic strength of the cues in each condition. In addition to possible differences in the inherent acoustic strength of cues in the different intonation conditions, a tendency for listeners whose native language is English to associate stressed or accented syllables (often exhibiting a H tone accent) with the frequent pattern of English word-initial stress (Cutler & Carter, 1987) may have contributed to the bias to perceive final syllables with the H–L pattern as disyllabic words. However, even with the HL Final Intonation pattern, the effect of distal prosodic pattern in the form of a reduction in disyllabic word reports in the monosyllabic Distal Context condition was observed.

5. General discussion

Three experiments investigated the question of which kinds of prosodic patterning in distal speech contexts can influence listeners' perception of prosodic structure. The mapping of prosodic structure onto a spoken utterance often results in the presence of repeated pitch patterns (Couper-Kuhlen, 1993; Dainora, 2001; Pierrehumbert, 2000), and/or of temporal patterns that give rise to perceptually equidistant stresses in time, i.e., perceptual isochrony (Lehiste, 1977). Previous examinations of distal prosody have shown that repeated patterns of tonal and temporal information can influence word segmentation in an ambiguous speech context (Dilley & McAuley, 2008). The current study investigated the extent to which a perceptual grouping hypothesis, as proposed by Dilley and McAuley (2008), can account for the perception of distinct prosodic structures in environments of varied acoustic correlates of prosodic constituency (f0 and duration cues), distinct intonation patterns, and tonal targets realized with different absolute fundamental frequencies. Experiments 1 and 2 used distal prosodic patterning consisting of typical “list” intonation (Beckman & Ayers Elam, 1997; Schubiger, 1958) to test whether this patterning would yield different perceived prosodic structures when the final two syllables had either (a) a LH tonal pattern (Experiment 1), or (b) a HL tonal pattern (Experiment 2). These experiments also tested whether prosodic patterning in the distal speech context involving f0 cues alone or durational cues alone can influence perceived prosodic structure of proximal syllables, and whether combined cues would yield stronger effects on perceived prosodic structure than either cue alone. Finally, Experiment 3 tested whether distal prosodic patterning consisting of “downtrends” – widely-described and common intonation patterns across languages that consist of gradually decreasing f0 values for H and L tones (Geluykens & Swerts, 1994; Lehiste, 1975; Streeter, 1978) – can influence the perception of proximal prosodic structure.

There are five key results from these experiments, each contributing to our understanding of the perception of prosodic structures in spoken language. First, the results of all three experiments clearly demonstrate that prosodic patterning in distal speech contexts strongly influences the perceived prosodic structure of proximal speech. When prosodic patterning in distal context set up an expectation for grouping of the final two syllables together, listeners heard a prosodic boundary corresponding to a Prosodic Word (or higher-level) prosodic constituent preceding those syllables and heard the final two syllables as grouped together into a single word. In contrast, when prosodic patterning in the distal context set up an expectation for the relatively large prosodic boundary to occur immediately before the final syllable, listeners heard the final syllable as its own group, thus perceiving it to be part of a separate prosodic constituent. This finding confirms the presence of distal prosodic effects similar to those found in a number of recent studies (Brown et al., 2011; Dilley et al., 2010; Dilley & McAuley, 2008) with different speech materials.

Second, the results of Experiments 1 and 2 clearly demonstrate that prosodic patterning based on a single cue, on either f0 alone or temporal manipulation alone, was capable of strongly influencing the perceived prosodic structure of proximal speech, with d' values ranging from 0.83 to 1.05 across those twos Cue Type conditions. Our experiments therefore show that prosodic patterning in distal speech context can influence perceived prosodic structure, even when that patterning is based on a single acoustic dimension.

A third key finding concerned the relative strengths of the effects of distal prosodic patterning on perception with cues from a single acoustic dimension, f0 alone, or duration alone, or with combined f0 and duration cues. The results of Experiments 1 and 2 were remarkably similar in showing that patterning based on duration cues alone or on f0 cues alone were equally effective in eliciting effects of distal prosodic patterning, while the combined f0 and duration cues showed the largest effects on perception of prosodic structure. This may be due to the fact that in English, naturally produced prosodic markers would often include both f0 and duration cues – either cue alone is not as effective in eliciting and effect on perception.

A fourth finding is that the effects of distal prosodic patterning on the perception of prosodic structure were not dependent on the specific intonation pattern of the final two syllables. All studies to date which have investigated contextual prosodic patterning effects on perception of lexico-prosodic structure have used a manipulation in which the final two syllables have had a LH pattern (Brown et al., 2011; Dilley et al., 2010; Dilley & McAuley, 2008). The results of both Experiments 2 and 3 clearly show that contextual prosodic patterning effects on proximal prosodic structure can also be obtained even when the final two syllables have a HL pattern.

A fifth finding is that the effects of distal prosody could be elicited in the environment of a “downtrend” intonation pattern. Such patterns are extremely common across languages and have been a topic of intensive phonetic study (Geluykens & Swerts, 1994; Lehiste, 1975; Streeter, 1978). Results of Experiment 3 reveal that distal prosodic patterns can influence the perception of proximal prosodic structure even if the f0 variation in the pattern does not consist of repeated *absolute f*0 values, i.e., when that patterning is a more abstract sequence of (phonologically defined) high and low pitches. The experiment therefore suggests that when f0 downtrend patterns occur in speech, prosodic expectations generated by the perception of repeated tonal groupings can still influence perception of prosodic boundaries in subsequent speech material.

In the following discussion, we first describe a number of implications of these findings for understanding the perception of prosody. The findings pertaining to prosody perception in turn have the implications for understanding the psycholinguistic process of word segmentation.

6. Implications for understanding perception of prosodic structure

The present results help to elucidate the nature of perception of prosodic structure. Most work on perception of prosodic structure has focused on the ways in which various kinds of proximal acoustic-phonetic variation determine perceived prosodic structures (Grabe, Kochanski, & Coleman, 2007; Hasegawa-Johnson et al. 2005; Streefkerk, Pols, & Ten Bosch, 1998). This prior work has largely assumed that the cues relevant to the perception of prosodic structure can be read off local (proximal) acoustic variation in attributes such as duration, f0, and intensity, e.g., as would be accomplished by an automatic speech recognizer.

The present work adds to a small but growing line of research showing that perception of prosodic structure is not merely a function of local acoustic information, but rather suggests that perception can be strongly influenced by prosodic information in the speech context. Previously, very little work had investigated the influence of prosodic context on speech perception and language processing. Existing work on this topic has largely focused on interpretation of ambiguous syntactic structures involving high vs. low attachment of words in a given syntactic tree as a function of prosodic context (Carlson, Clifton, & Frazier, 2001; Schafer, Speer, Warren, & White, 2000). These studies suggested that a given set of proximal (prosodic) word boundaries could be perceived as corresponding to prosodic phrase boundaries of different sizes (i.e., as an intermediate intonation phrase boundary or as an intonation phrase boundary). Recently, it has been claimed that prosodic boundaries, in particular, are utilized in resolving structural ambiguities by functioning as direct indicators of hierarchical structure (Carlson et al., 2009). However, while previous work has shown that prosodic boundaries serve as cues to hierarchical structure (Carlson et al., 2009), no explanatory mechanisms have been posited whereby prosodic context might have generated such effects.

The present work adds to recent findings showing that distal prosodic patterning in a speech context can influence the perception of proximal prosodic structure for a potentially large set of distinctive levels of the prosodic hierarchy. In particular, here and in other work (Brown et al., 2011; Dilley et al., 2010; Dilley & McAuley, 2008) a proximal juncture point consisting of fixed acoustic material could be heard as a relatively small, sublexical prosodic boundary (i.e., a *syllable edge*) – or the juncture point could be heard as the location of a relatively large prosodic boundary (i.e., of a *prosodic word* or larger unit). In other words, distal speech information influenced the interpretation of prosodic, and thus linguistic, structure. Although the stimuli in the current experiments consist of highly controlled realizations of prosodic patterning, the persistence of the distal prosodic effects with the incremental addition of more varied tonal patterns (such as downtrend) suggest that these effects could eventually be found in increasingly naturalistic speech environments. Distal speech contexts have also been shown to influence the perception of proximal acoustic material in the form of speech rate effects, which have been investigated in a separate line of work (Dilley & Pitt, 2010; Heffner, Dilley, McAuley, & Pitt, 2013). Presenting the distal speech at a slower rate resulted in listener reports that lacked the critical function word, even though the speech material containing the function word was acoustically identical.

The pattern of responses in our data suggests that distal prosodic patterning influenced not only the perception of prosodic boundaries, but also the perception of perceived prominence levels of syllables, e.g., whether a given syllable was heard as pitch accented. The lexical responses in the experimental tasks suggest that the final syllable of each experimental item (e.g., /bi/) was alternately perceived as the lexically unstressed syllable of a disyllabic word (e.g., *derby*) in the disyllabic Distal Context condition, or as a lexically stressed syllable bearing a pitch accent in the monosyllabic Distal Context condition (since monosyllabic content words are taken to be lexically stressed and able to bear a pitch accent (e.g., nip; see Pierrehumbert, 1980; Shattuck-Hufnagel & Turk, 1996). In other words, the final L or H tones syllable could be perceived as either accented or unaccented, depending on whether it was perceived to be the second (unaccented) syllable of a disyllabic word, or the first (accented) and only syllable of a monosyllabic word. Thus, we speculate that perceptual grouping affected not only the placement of prosodic boundaries, but also the relative prominence levels and locations of pitch accents within a prosodic constituent. These findings fit in with a larger body of research showing that the perception of prosodic prominence is generally influenced by the speech context. For example, repeating a word causes its degree of perceived prominence to decrease over time (Cole, Mo, & Hasegawa-Johnson, 2010). However, the type of phenomenon reported in Cole et al. (2010) is quite distinct from the phenomenon reported here, since the current results demonstrate that the perceived prominence of the same syllable can be altered by the distal context.

The present findings provide further support for the proposed perceptual grouping hypothesis (Dilley & McAuley, 2008) as a possible mechanism by which contextual effects of prosodic patterning influence perception of prosodic structure. General principles of auditory perceptual organization have been extensively investigated with non-speech auditory patterns such as tonal and musical sequences (Handel, 1989; McAuley, 2010; Povel & Essens, 1985), but the applicability of these principles in the domain of speech perception is less well-established. The experiments of Dilley and McAuley (2008) presented the first explicit evidence that implied perceptual grouping based on pitch and duration patterning could affect the perception of prosodic structure. Here, the perceptual grouping hypothesis successfully predicted effects of distal prosodic patterning on proximal prosodic structure with independent manipulations of f0 and duration cues with a novel set of lexical forms, and with a diverse set of intonation patterns – not only LH-final intonation pattern (Experiment 1), but also HL-final intonation (Experiment 2), as well as downtrend patterning (Experiment 3). Overall, the influence of distal prosodic context on both grouping (i.e., prosodic constituency) and meter (i.e., prosodic prominence) demonstrated here is consistent with the proposed perceptual grouping hypothesis. These results therefore suggest that perceptual grouping effects could extend to more naturalistic speech environments, where the realization of prosodic domains encompasses a broad range of varying acoustic cues. Continuing to increase the naturalness of the linguistic contexts in which these effects are found is one of the goals of future research in this area.

The proposal for a role for perceptual grouping in speech processing also provides a promising contribution to the understanding of the complex relationship between prosodic structure and quantifiable measures of linguistic rhythm. It is now widely acknowledged

that the traditional division of languages into distinct rhythmic classes is problematic, since many metrics, such as those based on segment duration and quality, do not correspond to "rhythm" as established in auditory perception, or in theoretical accounts of cross-linguistic prosodic structures (Hyman, 2006; Ramus, Nespor, & Mehler, 1999; White & Mattys, 2007). Arvaniti (2009) has argued that rhythm-based effects such as those first reported in Dilley and McAuley (2008) and further investigated in the current studies, present a productive means of investigating rhythmic structure cross-linguistically. A perceptually-motivated explanation for the organization of acoustic input into prosodic constituents may account for the consistent perception of linguistic rhythm in the face of large amount of phonetic and acoustic variability.

The present work also has implications for understanding constraints on possible structure in intonational phonology. According to Pierrehumbert (2000), proposals within the autosegmental-metrical (AM) theory of intonation (e.g., Beckman & Pierrehumbert, 1986; Pierrehumbert, 1980) have not adequately dealt with facts about intonational repetition, namely, that of the total of 216 theoretically possible accentual combinations, only a fraction are actually produced in natural speech (Pierrehumbert, 2000). On the other hand, well-articulated *parallelism principles* applied to non-speech tonal perception (Lerdahl & Jackendoff, 1983) state that sequences of musical tones (notes) will be grouped perceptually so that they are arranged in a pattern that contains parallel metrical structures across groups (i.e., parallel prominence patterns). These principles would constrain perceptual organization of meter, prominence, and grouping/phrasing, thereby limiting the set of possible structures that occur in music. Building on Lerdahl and Jackendoff, we propose that the capacity of speakers to produce combinations of low and high phonological tone sequences is constrained by perception, such that speakers tend to produce sequences of prominences and boundaries which obey the parallelism principle (i.e., sequences which listeners will hear as parallel). The current results serve as an initial investigation into the extent to which certain variations in the realization in prosodic structures are nevertheless perceived as parallel, in the sense that sequences of syllables are perceived as grouped together in a repeating pattern. The limited set of phrasal structures which occur in sequence (Pierrehumbert, 2000) may therefore be explained in terms of constraints on the perceptual capacities of speakers and listeners to hear repeated sequences of phonological tones as having similar phrasal constituency and prominence structure. While the exact degree to which parallel structures are perceived or realized in natural speech remains an open question, the effects of perceptual grouping on lexical perception provide a promising avenue by which to investigate this possibility.

7. Summary

In summary, these results show that many types of prosodic patterning in distal speech context strongly influence the perceived prosodic structure of proximal speech. Distinct distal prosodic cues, realized with f0 and duration manipulations, elicit effects both independently and when combined. The combined cues are stronger at inducing different proximal prosodic structures. Moreover, the effects of distal prosody on perception occur in the environment of intonational downtrends, which provides the strongest evidence thus far for the potential of distal prosodic effects to be common in everyday speech perception. Because the phonetic cues to prosodic structure vary greatly in spoken language, and, in particular, do not consist of tonal targets realized as absolute fundamental frequencies, the current results are crucial to demonstrating that the cues associated with perceptual grouping are available to listeners, and could routinely influence their perceptions of prosodic boundaries, thus affecting lexical perception. These results show that listeners' organization of speech into prosodic structures is a function of the perceptual tendency to organize syllable sequences into rhythmic (i.e., metrical and grouping) structures, and that such rhythmic organizations are strongly influenced by prosodic patterning in distal speech context.

Acknowledgements

We would like to thank Editor Martine Grice and two anonymous reviewers for their thoughtful comments, which greatly improved the manuscript. We thank Prashanth Rajarajan, Krista Bur, Evamarie Burnham, and Elizabeth Wieland for help with the experiment, and also thank Leif and Madeleine McAuley. This research was partially supported by NSF CAREER Award BCS 0874653 to L. Dilley.

Appendix. Syllable sequences

1. banker helpful (tie murder bee/timer derby)
2. kettle heaven (Tim burrow bow/timber oboe)
3. pebble dollar (bar lever chew/barley virtue)
4. gossip oyster (pan treaty coy/pantry decoy)
5. plenty fluid (tray dirty crease/traitor decrease)
6. angry index (lay birdie fence/labor defense)
7. feather onion (bay beaker few/baby curfew)
8. chapter elbow (rue beaver gin/ruby virgin)
9. magic notice (gang sterling go/gangster lingo)
10. kitchen dealer (may beanie grow/maybe negro)

11. hero vacuum (sell early gull/cellar legal)
12. bullet junior (come feeding key/comfy dinky)
13. liquid perish (broad leasing king/broadly sinking)
14. lumpy danger (chair eager knee/cherry gurney)
15. lender dentist (hare umber lap/harem burlap)
16. plasma honey (pigs typo low/pigsty polo)
17. forest pepper (pee canter might/pecan termite)
18. blanket mounted (ham mercy nick/hammer scenic)
19. magnet guilty (cry sister nip/crisis turnip)
20. tourist robin (draw musty plea/drama steeply)
21. sandwich rosy (far gopher meant/Fargo ferment)
22. trouble wealthy (limb burner sing/limber nursing)
23. nicely equal (gray veto stir/gravy toaster)
24. nature lazy (faux meaty tour/foamy detour)
25. lady jacket (bran diesel tree/brandy sultry)
26. fever pencil (lie bully word/libel leeward)
27. husband lemon (fan seaman cheese/fancy munchies)
28. fortune decade (win deeper fume/windy perfume)
29. center northern (two cancer plus/toucan surplus)
30. mixture pleasure (class seedy pose/classy depots)

References

Akaike, H. (1973). Information theory as an extension of the maximum likelihood principle in Proceedings of the Second International Symposium on Information Theory, B. N. Petrov, and F. Csaki, (Eds.) Pages 267–281.

Arvaniti, A. (2009). Rhythm, timing and the timing of rhythm. *Phonetica*, *66*, 46–63.

Ashby, M. (1978). A study of two English nuclear tones. *Language and Speech*, *21*, 326–336.

Bates, D., Maechler, M., & Bolker, B. (2012). lme4: Linear mixed-effects models using S4 classes. R package version 0.999375-42.

Beckman, M., & Ayers Elam, G. (1997). Guidelines for ToBI labeling, version 3. Ohio State University.

Beckman, M., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Boersma, P., & Weenink, D. (2002). Praat: Doing phonetics by computer [Computer program] (4.0.26 ed.). Software and manual available online at ⟨http://www.praat.org⟩.

Boltz, M. G. (1993). The generation of temporal and melodic expectancies during musical listening. *Perception and Psychophysics*, *53*, 585–600.

Brown, M., Salverda, A. P., Dilley, L. C., & Tanenhaus, M. K. (2011). Expectations from preceding prosody influence segmentation in online sentence processing. *Psychonomic Bulletin and Review*, *18*, 1189–1196.

Carlson, K., Clifton, C. J., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*, 58–81.

Carlson, K., Frazier, L., & Clifton, C. J. (2009). How prosody constrains comprehension: A limited effect of prosodic packaging. *Lingua*, *119*, 1066–1082.

Cho, T., McQueen, J., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics*, *35*, 210–243.

Christophe, A., Gout, A., Peperkamp, S., & Morgan, J. (2003). Discovering words in the continuous speech stream: The role of prosody. *Journal of Phonetics*, *31*, 585–598.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, *51*, 523–547.

Cohen, A., & 't Hart, J. (1968). On the anatomy of intonation. *Lingua*, *19*, 177–192.

Cohen, A., Collier, R., & 't Hart, J. (1982). Declination: construct or intrinsic feature of speech pitch? *Phonetica*, *39*, 254–273.

Cole, J., Mo, Y., & Hasegawa-Johnson, M. (2010). Signal-based and expectation-based factors in the perception of prosodic prominence. *Laboratory Phonology*, *1*, 425–452.

Couper-Kuhlen, E. (1993). *English speech rhythm: Form and function in everyday verbal interaction*. Amsterdam, Netherlands: John Benjamins.

Crystal, D. (1969). *Prosodic systems and intonation in English*. Cambridge: Cambridge University Press.

Cutler, A. (1990). Exploiting prosodic probabilites in speech segmentation. *Cognitive models of speech processing*.

Cutler, A., & Butterfield, S. (1992). Rhythmic cues to speech segmentation: Evidence from juncture misperception. *Journal of Memory and Language*, *31*, 218–236.

Cutler, A., & Carter, D. M. (1987). The predominance of strong initial syllables in the English vocabulary. *Computer Speech and Language*, *2*, 133–142.

Cutler, A., Dahan, D., & van Donselaar, W. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech*, *40*, 141–201.

D'Imperio, M. (2000). *The role of perception in defining tonal targets and their alignment* (Unpublished Ph.D. dissertation). The Ohio State University.

Dainora, A. (2001). *An empirically based probabilistic model of intonation in American English*. Chicago, Illinois: University of Chicago.

Dilley, L. C., Mattys, S., & Vinke, L. (2010). Potent prosody: Comparing the effects of distal prosody, proximal prosody, and semantic context on word segmentation. *Journal of Memory and Language*, *63*, 274–294.

Dilley, L. C., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 294–311.

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

Fiorentino, R., & Poeppel, D. (2007). Compound words and structure in the lexicon. *Language and Cognitive Processes*, *22*, 953–1000.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Gee, J. P., & Grosjean, F. (1983). Performance structures: A psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458.

Geluykens, R., & Swerts, M. (1994). Prosodic cues to discourse boundaries in experimental dialogues. *Speech Communication*, *15*, 69–77.

Grabe, E., Kochanski, G., & Coleman, J. (2007). Connecting intonation labels to mathematical descriptions of fundamental frequency. *Language and Speech*, *50*, 281–310.

Grice, M. (1995). Leading tones and downstep in English. *Phonology*, *12*, 183–233.

Handel, S. (1989). *Listening: An introduction to the perception of auditory events*. Cambridge, MA: MIT Press.

Hasegawa-Johnson, M., Chen, K., Cole, J., Borys, S., Kim, S.-S., Cohen, A., Zhang, T., Choi, J.-Y., Kim, H., Yoon, T., & Chavarria, S. (2005). Simultaneous recognition of words and prosody in the Boston UNiversity Radio Speech Corpus. *Speech Communication*, *46*, 418–439.

Hayes, B., & Lahiri, A. (1991). Bengali intonational phonology. *Natural Language and Linguistic Theory*, *9*, 47–96.

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*, 1275–1302.

Hyman, L. H. (2006). Word-prosodic typology. *Phonology*, *23*.

Jaeger, T. F. (2008). Categorical data analysis: Away from ANOVAs (transformation or not) and towards logit mixed models. *Journal of Memory and Language*, *59*, 434–446.

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review*, *83*, 323–355.

Jones, M. R., & Boltz, M. G. (1989). Dynamic attending and responses to time. *Psychological Review*, *96*, 459–491.

Jones, M. R., Boltz, M. G., & Kidd, G. (1982). Controlled attending as a function of melodic and temporal context. *Perception and Psychophysics*, *32*, 211–218.

Jones, M. R., Moynihan, H., MacKenzie, N., & Puente, J. (2002). Temporal aspects of stimulus-driven attending in dynamic arrays. *Psychological Science*, *13*, 313–319.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Columbus, OH: The Ohio State University.

Kim, S. (2004). *The role of prosodic phrasing in Korean word segmentation* (Unpublished dissertation). Los Angeles: Univeristy of California.
Knight, R. -A. (2003). *Peaks and plateaux: The production and perception of intonational high targets in English* (Doctoral dissertation). *University of Cambridge*.
Kubozono, H. (1989). Syntactic and rhythmic effects on downstep in Japanese. *Phonology*, *6*, 39–67.
Ladd, D. R. (2008). *Intonational phonology* (2nd ed.). Cambridge: Cambridge University Press.
Larkey, L. S. (1983). Reiterant speech: An acoustic and perceptual validation. *Journal of the Acoustical Society of America*, *73*, 1337–1345.
Lehiste, I. (1975). The phonetic structure of paragraphs. In: A. Cohen, & S. G. Nooteboom (Eds.), *Structure and Process in Speech Perception* (pp. 195–206). Berlin: Springer Verlag.
Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–263.
Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music*. Cambridge, MA: MIT Press.
Liberman, M., & Pierrehumbert, J. (1984). Intonational invariance under changes in pitch range and length. In: M. Aronoff, & R. Oerhle (Eds.), *Language sound structure* (pp. 157–233). Cambridge, MA: MIT Press.
MacMillan, N. A., & Creelman, C. D. (1991). *Detection theory: A user's guide*. New York: Cambridge University Press.
Mattys, S. L., & Melhorn, J. F. (2007). Sentential, lexical, and acoustic effects on the perception of word boundaries. *Journal of the Acoustical Society of America*, *122*, 554–567.
McAuley, J. D. (2010). Tempo and rhythm. *Music Perception*, 165–199.
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 & Performance*, *29*, 1102–1125.
Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the on-line syntactic analysis of spoken sentences. *Journal of Experimental Psychology: Learning, Memory and Cognition*, *34*, 874–885.
Nespor, M., & Vogel, I. (2007). *Prosodic phonology*. Berlin: Mouton De Gruyter.
Ohala, J. J. (1975). The temporal regulation of speech. In: Gunnar Fant, & M. A.A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 431–453).
Pierrehumbert, J. (1979). The perception of fundamental frequency declination. *Journal of the Acoustical Society of America*, *66*, 363–369.
Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation* (Unpublished Ph.D. dissertation). Cambridge, MA: MIT.
Pierrehumbert, J. (2000). Tonal elements and their alignment. In: M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 11–36). Dordrecht: Kluwer Academic Publishers.
Pierrehumbert, J., & Beckman, M. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
Pike, K. L. (1945). *The intonation of American English*. Ann Arbor: University of Michigan Publications.
Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception*, *2*, 411–440.
Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics*, *23*, 429–451.
Pynte, J., & Prieur, B. (1996). Prosodic breaks and attachment decisions in sentence parsing. *Language and Cognitive Processes*, *11*, 165–191.
Ramus, F., Nespor, M., & Mehler, J. (1999). Correlates of linguistic rhythm in the speech signal. *Cognition*, *73*, 256–292.
Salverda, A. P., Dahan, D., & McQueen, J. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.
Schafer, A. J., Speer, S. R., Warren, P., & White, S. D. (2000). Intonational disambiguation in sentence production and comprehension. *Journal of Psycholinguistic Research*, *29*, 169–182.
Schaffer, D. (1984). The role of intonation as a cue to topic management in conversation. *Journal of Phonetics*, *12*, 327–344.
Schubiger, M. (1958). *English intonation, its form and function*. Tübingen: Max Niemeyer.
Selkirk, E. O. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge, MA: MIT Press.
Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.
Streefkerk, B. M., Pols, L. C. W., & Ten Bosch, L. F. M. (1998). Automatic detection of prominence (as defined by listeners' judgments) in read aloud Dutch sentences. In Proceedings of ICSLP 1998 (Sydney). 3, 683–686.
Streeter, L. A. (1978). Acoustic determinants of phrase boundary perception. *Journal of the Acoustical Society of America*, *64*, 1582–1592.
Swerts, M., & Geluykens, R. (1994). Prosody as a marker of information flow in spoken discourse. *Language and Speech*, *37*, 21–43.
Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, *71*, 1596–1605.
Tilsen, S. (2012). Utterance preparation and Stress Clash: Planning prosodic alternations. In: S. Fuchs, P. Perrier, M. Weirich, & D. Pape (Eds.), *Speech production and perception: Planning and dynamics*.
Truckenbrodt, H. (2001). Downstep, upstep, and register levels. *Phonology Circle*, 1–16.
Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397–440.
van den Berg, R., Gussenhoven, C., & Rietveld, T. (1992). Downstep in Dutch: Implications for a model. In: G. J. Docherty, & D. R. Ladd (Eds.), *Gesture, segment, prosody* (pp. 335–367). Cambridge University Press, http://dx.doi.org/10.1017/CBO9780511519918.015.
Wagner, M. (2010). Prosody and recursion in coordinate structures and beyond. *Natural Language & Linguistic Theory*, *28*, 183–237.
Watson, D., & Gibson, E. (2004). The relationship between intonational phrasing and syntactic structure in language production. *Language and Cognitive Processes*, *19*, 713–755.
Welby, P. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation*. Columbus, OH: The Ohio State University.
Wheeldon, L., & Lahiri, A. (1997). Prosodic units in speech production. *Journal of Memory and Language*, *37*, 356–381.
White, Laurence, & Sven L., Mattys (2007). Calibrating rhythm: First language and second language studies. *Journal of Phonetics*, *35*(4), 501–522.
Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, *91*, 1707–1717.