Journal of Memory and Language 59 (2008) 294–311

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Distal prosodic context affects word segmentation and lexical processing ☆

Laura C. Dilley [a,b,*], J. Devin McAuley [a]

[a] Department of Psychology, Bowling Green State University, 200 Psychology Building, Bowling Green, OH 43403, USA
[b] Department of Communication Disorders, Bowling Green State University, 247 Health Center, Bowling Green, OH 43403, USA

ARTICLE INFO

Article history:
Received 11 December 2007
Revision received 20 June 2008
Available online 15 August 2008

Keywords:
Prosody
Word segmentation
Lexical access
Word recognition
Rhythm
Perceptual organization

ABSTRACT

Three experiments investigated the role of distal (i.e., nonlocal) prosody in word segmentation and lexical processing. In Experiment 1, prosodic characteristics of the initial five syllables of eight-syllable sequences were manipulated; the final portions of these sequences were lexically ambiguous (e.g., *note bookworm, notebook worm*). Distal prosodic context affected the rate with which participants heard disyllabic final words, although identical acoustic material was judged. In Experiment 2, removing four syllables of initial context reduced the magnitude of the distal prosodic effect. Experiment 3 used a study-test recognition design; better recognition was demonstrated for visually-presented disyllabic words when these items were comprised of adjacent syllables previously heard in distal prosodic contexts predicted to facilitate perceptual grouping of these two syllables. Overall, this research identifies distal prosody as a new factor in word segmentation and lexical processing and provides support for a perceptual grouping hypothesis derived from principles of auditory perceptual organization.

Introduction

A central problem in the study of spoken language concerns how the acoustic speech signal is mapped to words by the listener. A listener must both segment the continuous acoustic signal into candidate words, as well as perform lexical access by comparing the incoming acoustic material with stored lexical units in memory. How these processes are carried out constitute nontrivial problems, in part because word boundaries are not consistently marked by silences or any other acoustic cues (Cole & Jakimik, 1980; Klatt, 1980).

There have been a number of theoretical accounts of how word segmentation and lexical access take place (Marslen-Wilson & Welsh, 1978; McClelland & Elman, 1986; Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997). A critical issue for all such accounts is how different types of acoustic information influence both word segmentation and identification and selection of word candidates. In this regard, an important theoretical issue concerns the role of suprasegmental, as opposed to segmental, acoustic information in spoken language processing.

Prior research on the role of suprasegmental information in word segmentation and lexical access has focused almost entirely on contributions of suprasegmental cues that are *proximal* to (i.e., at or immediately adjacent to) the point where segmentation or lexical access of a word occurs. Lexical stress is one example of a proximal cue. Lexically stressed syllables differ from unstressed syllables in both segmental and suprasegmental properties. With

☆ Portions of this research were presented at the 10th Laboratory Phonology Conference and at the 47th Annual Meeting of the Psychonomic Society. The authors are especially grateful to Sven Mattys, Alice Turk and an anonymous reviewer for many helpful comments, suggestions, and discussions concerning this work. We are also grateful to Stefanie Shattuck-Hufnagel for many useful discussions as well as collaboration on pilot work leading up to the present paper. We also thank Anne Pier Salverda, Mari Riess Jones, Janet Pierrehumbert, Oliver Niebuhr, Heinz Giegerich, Sun-Ah Jun, Jacqueline Vassière, Cecile Fougeron, and Laurence White for helpful feedback and suggestions; Molly Henry for assistance with manuscript revisions and other useful input; Louis Vinke for help with data collection; and Mary Hare, Madeleine McAuley, and the members of the Rhythm, Attention, and Perception (RAP) Lab at Bowling Green State University for their helpful input and suggestions during the completion of this project.

* Corresponding author. Fax: +1 419 372 6013.
E-mail address: dilley@bgsu.edu (L.C. Dilley).

0749-596X/$ - see front matter © 2008 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2008.06.006

respect to suprasegmental properties, stressed syllables typically have longer duration, higher pitch, more uniform spectral balance, and/or higher amplitude than unstressed syllables; these properties also apparently distinguish levels of lexical stress, e.g., primary vs. secondary (Fry, 1955, 1958; Lehiste, 1970; Mattys, 2000; Sluijter & van Heuven, 1996).

Prior research has shown that listeners not only perceive proximal suprasegmental differences among syllables with similar segmental (i.e., full vowel) quality (Mattys, 2000; Mattys, 2004), but they are capable of using proximal suprasegmental cues in word segmentation (Banel & Bacri, 1994; Mattys, Jusczyk, Luce, & Morgan, 1999; Morgan, 1996; Nakatani & Schaffer, 1978; Quené, 1992, 1993; Vroomen & de Gelder, 1997; Vroomen, Tuomainen, & de Gelder, 1998). Moreover, recent evidence suggests proximal prosodic (e.g., fundamental frequency and/or durational) cues affect lexical access and competition (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Cutler & Donselaar, 2001; Davis, Marslen-Wilson, & Gaskell, 2002; Gout, Christophe, & Morgan, 2004; Shatzman & McQueen, 2006; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Salverda, Dahan, Tanenhaus, Crosswhite, Masharov, & McDonough, 2007; Millotte, René, Wales, & Christophe, 2008).

Of particular relevance for the present paper is recent work examining effects of proximal prosodic constituents on linguistic processing (Cho, McQueen, & Cox, 2007; Christophe et al., 2004; Gout et al., 2004; Millotte et al., 2008). These studies have shown that processing of lexical and/or syntactic ambiguities is influenced by the size of a prosodic boundary in the vicinity of the ambiguity; prosodic boundary size in these studies was defined according to the theory of the prosodic hierarchy, which proposes that segmental material is hierarchically organized into prosodic constituents of various sizes (e.g., Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; see Shattuck-Hufnagel & Turk, 1996, for a review). For example, Christophe et al. (2004) used stimuli in which a target syllable (e.g., [ʃa]) could form the beginning of a disyllabic word (e.g., *chagrin* 'sorrow') or else a monosyllabic word (e.g., *chat* 'cat', as in the phrase *chat grincheux* 'grumpy cat'). Recognition of the disyllabic word (e.g., *chagrin*) was delayed relative to a matched control phrase containing no potential lexical competitor when the target syllable (e.g., [ʃa]) was realized with a boundary ending a relatively small constituent (i.e., a prosodic word boundary). In contrast, recognition of the disyllabic word was not delayed relative to the matched control phrase when the target syllable was realized with a relatively larger prosodic boundary (i.e., a phonological phrase boundary).[1] This was signaled acoustically by proximal durational and/or fundamental frequency differences associated with the two different types of prosodic phrase boundaries.

The aim of this article is to investigate the hypothesis that more *distal* (i.e., distant or nonlocal) prosodic characteristics of spoken language affect perceived relative strengths of *proximal* prosodic boundaries, thereby influencing word segmentation and lexical processing. As far as we are aware, no published work has examined this issue. Nonetheless, there is evidence suggesting that distal prosodic cues influence spoken language processing. For example, distal prosody can affect the speed of phoneme monitoring (Cutler, 1976; Meltzer, Martin, Mills, Imhoff, & Zohar, 1976; Pitt & Samuel, 1990; Quené & Port, 2005; Shields, McHugh, & Martin, 1974), as well as the location of category boundaries in voice onset time continua (e.g., Kidd, 1989). Moreover, the interpretation of ambiguous syntactic structure is apparently influenced by the relative sizes of prosodic boundaries in a sentence (Carlson, Clifton, & Frazier, 2001; Schafer, Speer, Warren, & White, 2000); in addition, the processing of words which represent potential lexical embeddings (e.g., *cap* vs. *captain*) changes, depending on the position within prosodic structure (Salverda et al., 2007). Such findings suggest, but do not definitively show, that processing of proximal prosodic characteristics may be influenced by more distal ones.

Motivation for an effect of distal prosody on word segmentation and lexical processing comes from a relatively large literature on non-speech auditory perception illustrating effects of frequency, duration, and amplitude patterning on perceived organization of auditory sequences; see Handel (1989) for a review. In general, when individuals hear simple tone sequences, the frequency, duration, and amplitude patterning of sequence elements (i.e., tones) conveys a sense of sequence organization and structure. Perceived organization includes the sense that some sequence elements belong together (i.e., they are grouped), that within a group some elements are accented, while others are not, and that accent patterns tend to repeat. For example, in an isochronous sequence of tones of equal amplitude and duration alternating between a fixed high (H) and fixed low (L) frequency, e.g., HLHLHL, listeners tend to hear a repeating strong–weak binary grouping of tones with either the high or low tone as accented and beginning the group, i.e., (H̲L)(H̲L)(H̲L) or (L̲H)(L̲H)(L̲H) (Woodrow, 1909, 1911).

Three key findings in this literature form the basis for the present investigation. First, repeating strong–weak binary patterns of accents induced by distal frequency, duration, and/or amplitude patterning of sequence elements tend to generate periodic expectations about the grouping and perceived accentuation of later sequence elements, even when there are no explicit proximal acoustic cues to grouping and accents in those elements (Boltz, 1993; Jones, 1976; Jones & Boltz, 1989; Large & Jones, 1999; McAuley & Jones, 2003; Povel & Essens, 1985; Thomassen, 1982). Second, repeating accent patterns induced by distal frequency and/or timing characteristics tend to produce stronger expectations with more pattern repetitions (i.e., when there is more distal context) (Bregman, 1978; Jones & Yee, 1993; McAuley & Kidd, 1998). Third, periodic expectations induced by distal pattern structure do not re-

[1] A prosodic word is a roughly word-sized constituent containing only one lexical head potentially grouped with one or more functional elements (Nespor & Vogel, 1986; see Shattuck-Hufnagel & Turk, 1996, for a review). A phonological phrase is a unit containing one or more prosodic words that typically line up with syntactic constituents and are marked by pre-boundary lengthening (Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992).

quire explicit stimulus markers in order to persist to influence the processing of later sequence elements; thus, a missing (silent) element that is introduced into a rhythmic sequence is perceived as an unaccented element or 'silent beat' in both music and language (Large & Jones, 1999; Large & Palmer, 2002; McAuley & Kidd, 1998; Parncutt, 1994; Povel & Essens, 1985; Selkirk, 1984).

Based on these findings, we propose a *perceptual grouping hypothesis*, namely that prosodic (i.e., fundamental frequency and duration) cues distal from the locus of segmentation or lexical access of a word affect the unfolding process of perceiving prosodic constituents, thereby influencing word segmentation and lexical recognition in a manner consistent with principles of perceptual organization for non-speech auditory patterns. The three key findings from the non-speech auditory perception literature lead to three corresponding predictions concerning potential effects of distal prosody on word segmentation and lexical processing: (1) distal F0 and duration patterns should impact the perceptual grouping of syllables into prosodic constituents, and therefore the lexical processing of candidate words indicated by the grouping, even when there are no *proximal* cues to this grouping; (2) distal F0 and duration patterns should have larger effects when there is more repetition; and (3) expectations about the perceptual grouping of syllables should persist in spite of a missing or 'silent' beat as long as the overall rhythmic patterning is maintained.

Prosodic repetition in pitch and rhythm as cued by F0 and duration are common in speech, suggesting that such regularities might be exploited in a number of listening situations. Indeed, similar proposals have been made by Martin and colleagues (Martin, 1972, 1979; Meltzer et al., 1976; Shields et al., 1974). With respect to pitch, it has widely been observed that prominence-lending pitch excursions tend to form repeating patterns of various types and sizes (Chafe, 1988; Crystal, 1969; Crystal & Quirk, 1964; Gibbon, 1976; Halliday, 1967; Kingdon, 1958; Ladd, 1986, 1996; Palmer, 1922; Pierrehumbert, 2000; Pike, 1945; Schubiger, 1958). Such repetition has been noted for 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; Prieto, van Santen, & Hirschberg, 1995; Welby, 2003). With respect to rhythm, the stresses of speech often sound perceptually isochronous (e.g., Couper-Kuhlen, 1993; Dilley, 1997; Lehiste, 1977; McAuley & Dilley, 2004).

These sorts of prosodic regularities could potentially provide distal cues that could influence proximal processing of lexical content. Very few other studies have explored distal effects of any kind on lexical processing. Two exceptions are studies by Gómez and colleagues (Gómez, 2002; Gómez & Maye, 2005), and Mattys, Melhorn, and White (2007), which examined the influence of distal statistical dependencies and distal syntactic expectations, respectively, on processing proximal lexical material.

Overview of experiments

In the present paper, we report three experiments which demonstrate effects of distal prosody on word segmentation and lexical processing. Eight-syllable target sequences were constructed, such as *channel dizzy foot note book worm,* where the final four syllables had ambiguous lexical structure and could be organized into words in more than one way (e.g., *footnote bookworm*, *foot notebook worm*). In all experiments, the final three syllables of each target sequence had fixed acoustics, while F0 and/or duration characteristics of the distal syllables were varied. For the remainder of the paper, the term 'proximal' will refer to speech syllables which comprise or are adjacent to either possible final lexical item—*worm* or *bookworm*—while 'distal' will refer to speech syllables preceding proximal material. Of interest here was a two-level distal prosodic context manipulation that was predicted to influence the perceptual grouping (i.e., the prosodic constituency) of the final syllables in the target sequence. In particular, one of the prosodic contexts was predicted to cause a relatively stronger prosodic boundary to be heard before the penultimate syllable (e.g., *book*) than before the final syllable (e.g., *worm*), so that listeners would group the final two syllables as a single disyllabic word (e.g., *bookworm*); this context will be referred to as the Disyllabic context. In contrast, the other prosodic context was predicted to reverse the perceived relative strengths of the prosodic boundaries before the penultimate and final syllables, causing a stronger prosodic boundary to be heard before the final syllable (e.g., *worm*) than before the penultimate syllable, so that listeners would hear a final monosyllabic word (e.g., *worm*); this context will be referred to as the Monosyllabic context.

In an F0 condition, for both the Disyllabic and Monosyllabic contexts, the final (proximal) three syllables of each target sequence received a high (H), low (L), high (H) F0 pattern (one F0 target per syllable), and the repeating F0 pattern of the first five (distal context) syllables was varied. In the Disyllabic context, the first five syllables of each target sequence received a L_1-H_2-L_3-H_4-L_5- pattern, with one F0 target, H or L, on each syllable; here, subscripts indicate syllable numbers, while hyphens indicate syllable boundaries. In contrast, in the Monosyllabic context, the first five syllables of each target sequence received a H_1-L_2-H_3-L_4-HL_5- pattern with one F0 target for each of the first four syllables, and a fall in F0 from H to L on the fifth syllable. Because the final three syllables of each target sequence were consistently assigned a high (H), low (L), high (H) F0 pattern, the different repeating F0 patterns associated with the Disyllabic and Monosyllabic contexts were expected to alter the perceptual grouping (i.e., the prosodic constituency) of the final three syllables, and hence the segmentation of those syllables into words. Specifically, based on findings from the non-speech auditory perception literature concerning the perceptual organization of alternating high–low frequency patterns (e.g., Woodrow, 1909, 1911), the Disyllabic context was predicted to yield a (L_1-H_2-)(L_3-H_4-)(L_5-H_6-)(L_7-H_8) grouping of the target syllable sequence with a larger boundary before syllable 7 than syllable 8, thereby critically yielding a disyllabic final word report (e.g., *bookworm*). In contrast, the Monosyllabic context was predicted to yield a (H_1-L_2-)(H_3-L_4-)(HL_5-)(H_6-L_7-)(H_8...) grouping of the target sequence syllables with a larger boundary before syllable 8 than before sylla-

ble 7, critically yielding a monosyllabic final word report (e.g., *worm*).

In a Duration condition, we predicted that distal duration cues alone should affect the prosodic phrasing of a final three-syllable sequence with fixed acoustics, even when frequency cues are held constant at a monotone across the entire sequence. In particular, a distal context involving a periodic alternation of strong (S) and weak (W) syllables should cause listeners to continue to hear a binary, (SW) grouping of sequence elements even when there are no explicit cues to grouping in proximal material. For this manipulation, the Disyllabic context consisted of two trochaic, or SW, disyllabic words, plus a *relatively short* fifth syllable of about the same duration as the preceding syllables; the Monosyllabic context, in contrast, consisted of two SW disyllabic words, plus a *relatively long* fifth syllable of a duration approximately equal to that formed by the two preceding SW syllables. Based on a continuation of the alternating pattern of stresses, the Disyllabic context was predicted to yield a (S_1-W_2-)(S_3-W_4-)(S_5-W_6-)(S_7-W_8) perceptual grouping of the target sequence syllables and cause listeners to hear a stronger prosodic boundary before S_7 than before S_8 and thus report a disyllabic final word (e.g., *bookworm*). In contrast, lengthening the fifth syllable in the Monosyllabic context was expected to induce the sense of a 'silent beat', *w*, on the second-half of the lengthened syllable, and cause listeners to hear the lengthening on the fifth syllable as its own trochaic group (Sw_5-), resulting in a different perceptual grouping of target sequence syllables (S_1-W_2-)(S_3-W_4-)(Sw_5-)(S_6-W_7-)(S_8...). The shifted perceptual grouping of target sequence syllables introduced by the missing 'beat' was expected to move the location of the stronger prosodic boundary to before S_8 and thus cause listeners to report a monosyllabic final word (e.g., *worm*).

Finally, in a third F0+Duration condition, we combined the distal F0 and duration cues in a complementary fashion with the expectation that this would yield the strongest effects of distal prosody on perceived relative prosodic boundary strength, and hence the largest effects of distal prosody on segmentation and lexical processing, of the final syllables in the target sequences.

For both Experiments 1 and 2, participants listened to target and filler sequences and freely reported the last word they heard. The Monosyllabic context was expected to generate a stronger prosodic boundary before the final syllable than before the penultimate syllable and thus lead to monosyllabic final word reports, while the Disyllabic context was expected to generate a stronger prosodic boundary before the penultimate syllable than the final syllable and thus lead to disyllabic final word reports. Among the three distal prosody conditions, the largest effects were predicted for combined F0 and duration cues. Across Experiments 1 and 2, we varied the number of syllables of distal prosody with the expectation that distal prosody would have a greater impact on segmentation of longer sequences (Experiment 1) than on shorter, truncated sequences (Experiment 2), due to greater repetition. Experiment 3 extends the findings from the first two experiments to a study-test recognition design. If distal prosody affects the strength of lexical encoding of candidate words due to different prosodic constituent-induced segmentation patterns, then, in the context of a study-test recognition design, distal prosody should affect recognition accuracy for previously heard word candidates.

Experiment 1

Methods

Participants and design

One hundred and thirty-eight native speakers of American English completed the experiment in return for course credit in an introductory psychology course at the Ohio State University. Participants were at least 18 years old with self-reported normal hearing and a range a musical experience. The experiment implemented a 3 (type of prosody: F0, Duration, F0+Duration) $\times$ 2 (type of context: Disyllabic vs. Monosyllabic) mixed factorial design. Participants were randomly assigned to one of the three Type of Prosody conditions: F0 ($n = 57$), Duration ($n = 40$) or F0+Duration ($n = 41$). For each type of prosody, participants heard both Monosyllabic and Disyllabic distal contexts.

Materials

Twenty eight-syllable target sequences were constructed (see Appendix A). Each target sequence consisted of two disyllabic words with initial primary stress, e.g., *channel dizzy*, followed by a four-syllable string that could be organized into words in more than one way, e.g., *footnote-book-worm* can be organized as *footnote bookworm, foot notebook worm*, etc. Each of these final four syllables had full vowel quality and each could be a stressed, monosyllabic word. Moreover, 40 filler sequences were created ranging in length from 6 to 10 syllables. Fillers consisted of a mixture of monosyllabic and disyllabic words with unambiguous lexical structure in varying positions within the string; these were intended to disguise the lexical ambiguity present in target sequences. Disyllabic words in fillers always had initial primary stress; half of filler sequences ended in a monosyllabic word and half in a disyllabic word.

Target and filler sequences were read as connected speech by the first author, using monotone F0; the final four syllables of target sequences were spoken as two disyllabic words. Recordings were made in a quiet room onto DAT at a 16-kHz sampling rate using a Tascam DA-30MKII DAT recorder connected to an N/D308A cardioid microphone via a Yamaha MV802 mixer. Digitized utterances were then transferred from DAT to PC. A series of resynthesized speech stimuli were then derived from these utterances using the pitch-synchronous overlap-and-add (PSOLA) algorithm (Moulines & Charpentier, 1990) as implemented in Praat software (Boersma & Weenink, 2002).

F0 condition

The F0 characteristics of the initial five syllables were varied, while the temporal properties of these syllables were held constant. This was accomplished by creating stylized F0 contours during resynthesis consisting of a ser-

ies of local high (H) F0 maxima and low (L) F0 minima, connected by straight line interpolations. Values of H ranged from 270 to 280 Hz, while values of L ranged from 170 to 180 Hz. To create the Monosyllabic context within the F0 condition, a falling F0 pattern was generated across both of the initial two-syllable words; in addition, a falling F0 pattern was generated across the fifth syllable, which could be its own monosyllabic word. For example, in the eight-syllable target sequence *channel dizzy foot-note-book-worm*, *chan-*, *-nel*, *diz-*, and *-zy* were paired with H, L, H, and L, respectively, while *foot* was paired with a sequence of F0 targets, HL. (See Fig. 1, middle panel.) To create the Disyllabic context within the F0 condition, a rising F0 pattern was generated across both of the initial two-syllable words, while a *low* F0 was generated across the fifth syllable alone; for example, *chan-* was paired with L, *-nel* was paired with H, *diz-* was paired with L, *-zy* was paired with H, and *foot* was paired with L. (See Fig. 1, bottom panel.) Finally, for both Monosyllabic and Disyllabic contexts, the sixth, seventh, and eighth syllables were paired with H, L, and H targets, respectively. The resynthesis parameters for the final three syllables were held constant across both contexts.

Duration condition

For this condition, the temporal properties of prosodic contexts were varied while the F0 was held constant across each string. Praat was first used to resynthesize a single monotone version of each target string with F0 = 220 Hz. Monosyllabic and Disyllabic contexts were generated by waveform splicing in Praat to lengthen or shorten the interval from the vowel onset of the fifth syllable to the onset of the consonant or vowel in the sixth syllable; this interval will be referred to as the critical inter-onset-interval (IOI) (see Fig. 1). Consonant and vowel onsets were identified by inspecting waveforms and spectrograms for associated major amplitude discontinuities. To create the Monosyllabic context, the duration of the critical IOI was lengthened through splicing to approximately match the average duration of the IOI between vowel onsets of syllables 1 and 3 (e.g., *chan-* and *diz-* in *channel dizzy foot-note-book-worm*) and the IOI between syllables 3 and 5 (e.g., *diz-* and *foot* in *channel dizzy foot-note-book-worm*). For 17 of 20 target sequences in the Monosyllabic context, durational adjustments were made exclusively by duplicating portions of the fifth syllable nucleus and/or coda, and/or the closure duration of an initial stop consonant in the sixth syllable. For the remaining three of 20 target sequences, adjustments were also made to the onset consonant of the sixth syllable by duplicating portions of a fricative. The average amount of lengthening for critical IOIs in the Monosyllabic context was 162 ms. To create the Disyllabic context, the critical IOI was shortened through splicing to approximately match the average duration of the IOI between vowel onsets of syllables 6 and 7 (e.g., *note* and *book* in *channel dizzy foot-note-book-worm*) and syllables 7 and 8 (e.g., *book* and *worm* in *channel dizzy foot-note-book-worm*). For 15 of 20 target sequences, durational adjustments were made exclusively by splicing out portions of the fifth syllable nucleus and/or coda, and/or portions of the closure of an initial stop in the sixth syllable. For the remaining sequences, the duration of the initial consonant in the sixth

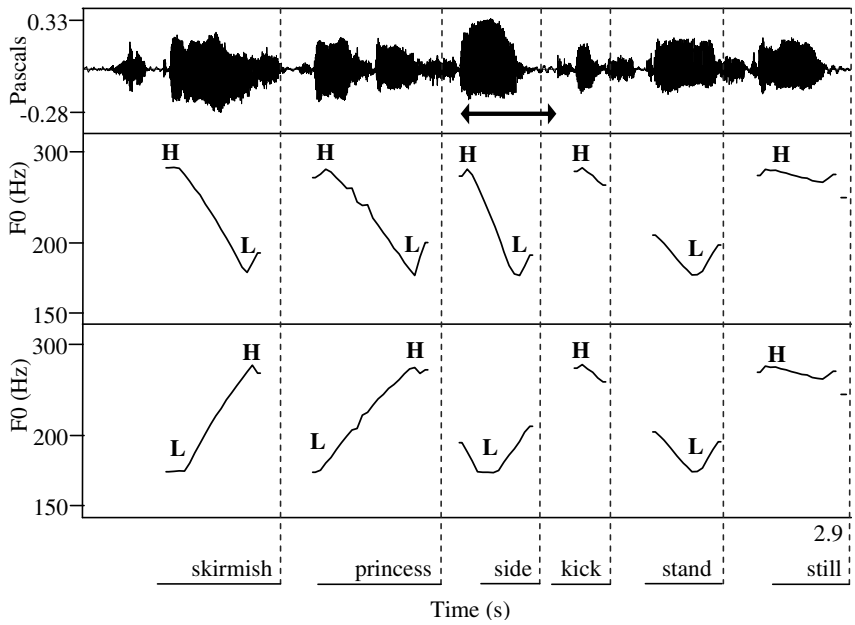

Fig. 1. Explanation of synthesis manipulation for target syllable sequences; shown is a sample stimulus from the F0 condition. The top panel shows a sample speech waveform. The bottom two panels show how the Monosyllabic context and Disyllabic context were created by varying the fundamental frequency (F0) contours across the initial five syllables of each target sequence; the acoustic characteristics of the final three syllables were held constant. The middle panel shows an example of a Monosyllabic context, and the bottom panel shows an example of a Disyllabic context. H and L refer to high and low F0 targets, respectively. The arrow indicates the portion of the speech signal which was lengthened or shortened in the Duration and F0+Duration conditions. See text for more information.

syllable was also subject to splicing. The average amount of shortening for critical IOIs in the Disyllabic context was 104 ms. For both types of context, all splices were made at zero crossings, and care was taken to ensure spectral continuity of formants when splicing e.g., from vowel midpoints. The naturalness of all stimuli was evaluated perceptually by the first author, who has extensive training in phonetics.

F0+Duration condition: Monosyllabic context stimuli from the F0 condition were subjected to the same splicing manipulations as Monosyllabic context stimuli from the Duration condition to create Monosyllabic context stimuli for this condition. Similarly, Disyllabic context stimuli from the F0 condition were subjected to the same splicing manipulations as Disyllabic context stimuli in the Duration condition to create Disyllabic context stimuli for this condition. F0 discontinuities arising through splicing were eliminated using the PSOLA algorithm in Praat by replacing any such points with gradual linear transitions; F0 contours were otherwise unaltered relative to the F0 condition.

Filler stimuli

For F0 and F0+Duration conditions, each filler sequence was resynthesized in two versions: as a series of repeated falling (HL) patterns and a series of repeated rising (LH) patterns; F0 values for H and L were in the range 270–280 and 170–180 Hz, respectively. Approximately half of fillers with a HL pattern ended in a word with falling F0, while the other half ended in a word with a final level high pitch, to ensure that experimental items were not the only sequences ending in a level high pitch. For the Rhythm Condition, fillers were resynthesized to have flat F0 of 220 Hz. There were thus 80 fillers for each type of prosody [F0 and F0+Duration conditions: 40 sequences $\times$ 2 F0 patterns (HL, LH) $\times$ 1 repetition; Duration condition: 40 sequences $\times$ 1 (flat) F0 pattern $\times$ 2 repetitions].

Procedure

Participants gave a free written report about the last word they heard in each sequence. Six filler sequences served as practice, followed by 20 target sequences and 80 fillers presented in a random order; each stimulus was followed by 2.3 s of silence. Half of target sequences for a participant were heard in a Monosyllabic context and the other half were heard in a Disyllabic context, with the specific target sequence–context pairing counterbalanced across participants. This resulted in two complementary lists for each type of distal prosody; two additional lists were constructed by reversing the orders of the first two lists, resulting in a total of four lists per prosody condition. Approximately equal numbers of participants were assigned to each list. The entire experiment took about 30 min to complete.

Results

Fig. 2A shows mean proportions of disyllabic final word responses with 95% confidence intervals as a function of type of context (Disyllabic vs. Monosyllabic) and type of prosody (F0, Duration, F0+Duration). A 2 (type of context) $\times$ 3 (type of prosody) mixed-measures ANOVA on disyllabic response proportions revealed a main effect of type of context ($F_1(1,135) = 286.26$, $MSE = 0.033$, $p < .001$; $F_2(1,19) = 111.91$, $MSE = 0.039$, $p < .001$; min-$F'(1,36) = 80.46$, $p < .001$), as well as a main effect of type of distal prosody ($F_1(2,135) = 3.988$, $MSE = 0.079$, $p < .05$; $F_2(2,38) = 8.341$, $MSE = 0.018$, $p < .01$; min$F'(2,162) = 2.70$, $p = .07$). There was also a significant interaction between type of prosody and type of context ($F_1(2,135) = 10.31$, $MSE = 0.033$, $p < .001$; $F_2(2,38) = 37.14$, $MSE = 0.005$, $p < .001$; min$F'(2,173) = 8.07$, $p < .001$). Consistent with the perceptual grouping hypothesis, there were significantly more disyllabic final word responses in the Disyllabic context condition than in the Monosyllabic context condition. The largest difference in response proportions between contexts was found for the F0+Duration condition ($M = 0.50$, 95% CI = 0.41–0.59), the next largest difference was found for the F0 condition ($M = 0.38$, 95% CI = 0.31–0.45), while the smallest difference was found for the Duration condition ($M = 0.24$, 95% CI = 0.18–0.30).

To consider the sensitivity of participants' final word reports to the distal context manipulation separately from possible response biases to report either monosyllabic or disyllabic final words, analyses of response proportions were supplemented by a signal detection analysis. Reporting a disyllabic final word in a Disyllabic context represented a response consistent with predictions of the perceptual grouping hypothesis; this was coded as a *hit*. Conversely, reporting a disyllabic final word in a Monosyllabic context represented a response which was inconsistent with the perceptual grouping hypothesis; this was coded as a *false alarm*. Hits and false alarms for each participant were then used to calculate the signal detection measures d' (a measure of sensitivity to type of distal context) and c (a measure of response bias).[2] Values of d' provided a standardized measure of the degree to which the type of final word report—disyllabic or monosyllabic—depended on whether the target sequence was paired with a Disyllabic and Monosyllabic distal context, respectively; $d' = 0$ indicated no effect of type of distal context on final word reports. In contrast, values of c provided a standardized measure of any bias in final word reports, 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.

Figs. 3A and B (gray bars) show mean values of d' and c with 95% confidence intervals for the F0, Duration, and F0+Duration conditions. A one-way between-subjects AN-

[2] Application of signal detection theory (SDT) provides a method to measure (index) sensitivity of final word reports to distal prosodic context. The most commonly used SDT measure of sensitivity is d', which is obtained by transforming participant hit rates (H) and false alarm rates (F) to z-scores and then calculating the difference: $d' = z(\mathrm{H}) - z(\mathrm{F})$. From a SDT standpoint, the bias measure, c, indexes participants' willingness to generate a particular response choice (i.e., degree of conservatism). It is defined as $c = -0.5[z(\mathrm{H}) - z(\mathrm{F})]$. In the context of the present study, c measured participants' willingness to report disyllabic final words, with values of c greater than zero indicating a more conservative response criterion (less willingness to report disyllabic words) and values less than zero indicating a more liberal response criterion (greater willingness to report disyllabic words). See MacMillan and Creelman (1991) for a comprehensive introduction to SDT.

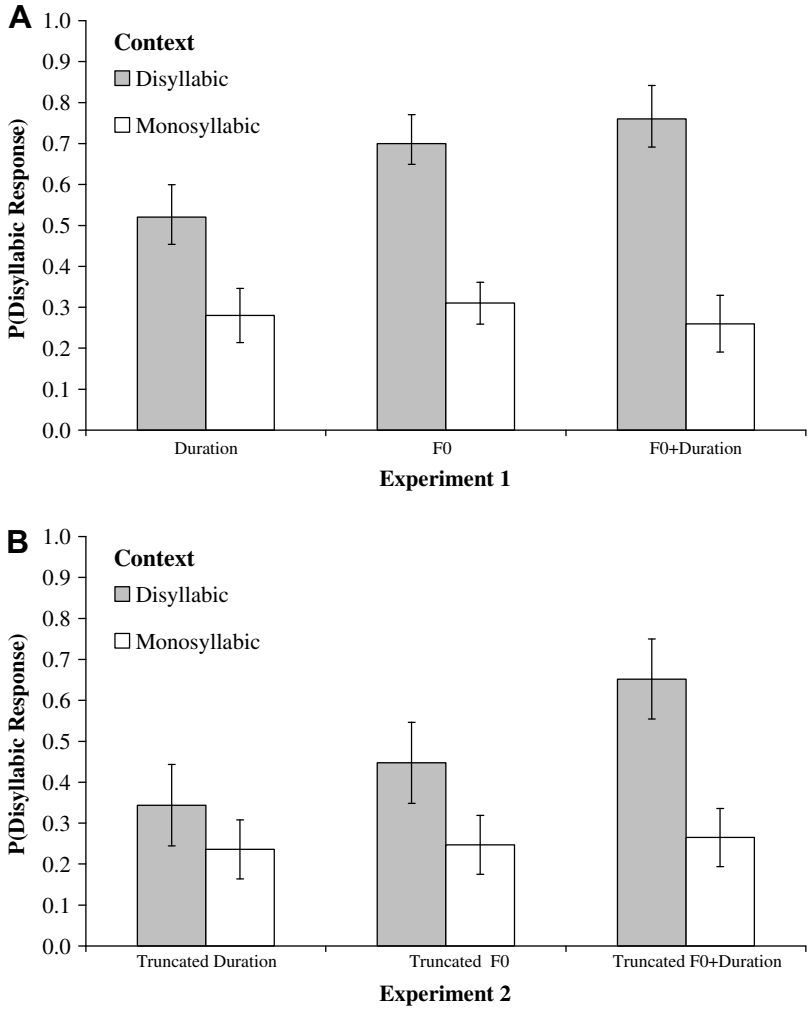

Fig. 2. Mean proportions of disyllabic final word responses with 95% confidence intervals for Disyllabic context and Monosyllabic context for the three prosodic conditions in Experiment 1 (A) and for the associated Truncated conditions in Experiment 2 (B).

OVA on d' revealed a main effect of type of prosody ($F_1(2,135) = 12.05$, $MSE = 0.664$, $p < .001$; $F_2(2,38) = 9.723$, $MSE = .258$, $p < .001$; $min\ F'(2,105) = 5.38$, $p < .01$). Consistent the analysis of disyllabic response proportions, d' scores were highest for the F0+Duration condition, next highest for the F0 condition, and lowest for the Duration condition; differences in d' for all pairs of conditions were reliable (F0 vs. Duration: $M = 0.48$, 95% CI = 0.08–0.87; F0+Duration vs. F0: $M = 0.41$, 95% CI = 0.02–0.80; F0+Duration vs. Duration: $M = 0.89$, 95% CI = 0.45–1.31). With respect to response bias, c, participants reported approximately equal numbers of monosyllabic and disyllabic final words for both F0 ($M = -0.01$, 95% CI = −0.18 to +0.15) and F0+Duration conditions ($M = -0.03$, 95% CI = −0.22 to +0.17). However, there was a slight tendency to report more monosyllabic words in the Duration condition ($M = 0.30$, 95% CI = 0.10–0.49). A one-way between-subjects ANOVA on c revealed a main effect of type of distal prosody for the subject analysis, but not for the item analysis ($F_1(2,135) = 3.50$, $MSE = 0.397$, $p < .05$; $F_2(2,38) = 0.585$, $MSE = 1.004$, $p = 0.56$; $\min F'(2,51) = 0.50$, $p = .61$). Post-hoc comparisons using Tukey's HSD found marginally reliable differences between the Duration condition and each of the two F0 cue conditions (F0, $p = 0.05$; F0+Duration, $p = .06$), but no difference for the F0 vs. F0+Duration comparison ($p = .99$).

Discussion

Overall, participants' final word reports are consistent with the view that participants formed different proximal patterns of prosodic constituency for lexically ambiguous syllable sequences based on both distal prosodic F0 and duration cues. Consistent with the perceptual grouping hypothesis, the distal Disyllabic context led to disyllabic final word reports, while the distal Monosyllabic context led to monosyllabic final word reports. Moreover, combining F0 and duration cues in a complementary fashion strengthened participants' perceived syllable groupings. For both the analysis of response proportions and d', the largest ef-

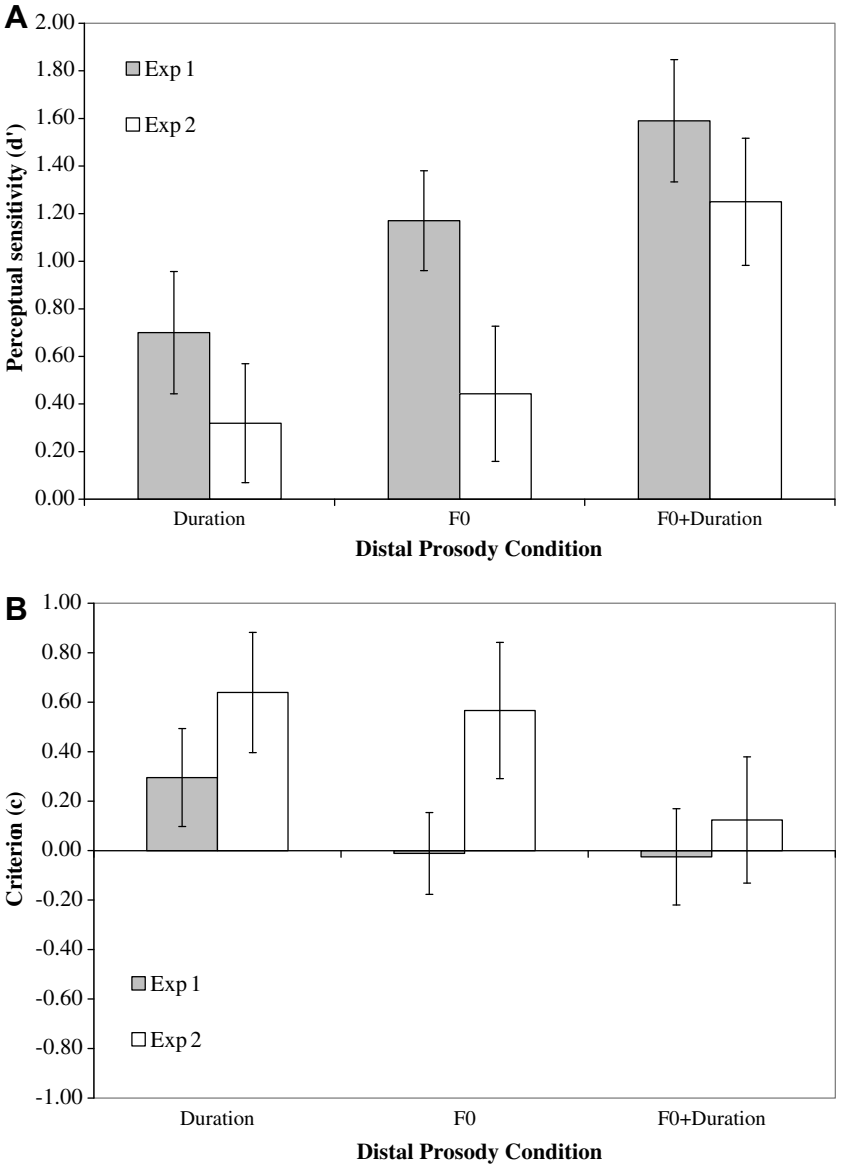

Fig. 3. Signal detection measures for Experiments 1 and 2. (A) Mean values of d' with 95% confidence intervals for the Duration, F0, and F0+Duration conditions in Experiment 1 (gray bars) and for the associated Truncated conditions in Experiment 2 (white bars); (B) Mean values of c with 95% confidence intervals for the Duration, F0, and F0+Duration conditions in Experiment 1 (gray bars) and for the associated Truncated conditions in Experiment 2 (white bars).

fect of distal context was found for the F0+Duration condition, the smallest effect was found for the Duration condition, and an intermediate effect was observed for the F0 condition. The stronger effect observed for the F0 condition relative to the Duration condition is consistent with other reports of perceptual dominance of pitch over durational cues in establishing stress or meter in language and music (Fry, 1958; Hannon, Snyder, Eerola, & Krumhansl, 2004; Spitzer, Liss, & Mattys, 2007).

One question raised by this work is that it is unclear precisely *how distal* the prosodic effects are. In particular, F0 and/or durational differences on the 5th syllable alone could conceivably have been responsible for parsing effects observed on the proximal syllables. For example, the durational lengthening and/or F0 drop on the 5th syllable in the Monosyllabic context across all three prosody conditions is expected to generate perception of a major prosodic phrase boundary at that location (cf. Turk & Sawusch, 1997; Turk & Shattuck-Hufnagel, 2000; Turk & White, 1999). Such changes in and of themselves might have caused listeners to hear a word boundary after this syllable. Subsequent syllables might then have been grouped so as to create the longest initial lexical candidate, such that e.g., *foot*–(PHRASE BOUNDARY)–*note–book–worm* was parsed as *foot notebook worm*. Another possibility is that the durational lengthening and/or F0 drop on the 5th syllable in Monosyl-

labic contexts could simply have caused this syllable (e.g., *foot*) to be a less viable onset of a two-syllable word (e.g., *footnote*); the 6th through 8th syllables (e.g., *note–book–worm*) could then have been grouped to create the longest initial lexical candidate, resulting in more monosyllabic final words (e.g., *worm*).

Both alternative explanations thus attribute the observed segmentation differences to manipulations in the vicinity of the 'near distal' 5th syllable, with no contribution of syllables 1–4. Moreover, neither explanation relies on the notion of perceptual grouping *per se* to explain proximal differences in perceived syllable groupings. In contrast, if the 'far distal' context of syllables 1–4 strengthens perceptual grouping of the final syllables by providing more instances of pattern repetition, then this should enhance context-dependent prosodic phrasing and segmentation of proximal material.

To contrast these alternative explanations with the 'far distal' account linked to the perceptual grouping hypothesis, Experiment 2 replicated Experiment 1 using shorter target sequences in which the initial four syllables were removed, thus reducing the amount of distal prosodic context. If the magnitude of the context-dependent segmentation effect on these truncated stimuli is the same as in Experiment 1, then it is likely that Experiment 1 results were due to 'near distal' effects unrelated to the perceptual grouping hypothesis, as suggested by the two alternative accounts. On the other hand, a reduced effect for the shorter (truncated) contexts compared with the longer (un-truncated) contexts examined in Experiment 1 will provide further support for the perceptual grouping hypothesis.

Experiment 2

Methods

Participants and design

One-hundred nine native speakers of American English completed the experiment in return for course credit at Ohio State University or Bowling Green State University. Participants were at least 18 years old with self-reported normal hearing and a range a musical experience. They were assigned to one of three prosodic conditions: Truncated F0 ($n = 36$), Truncated Duration ($n = 36$) or Truncated F0+Duration ($n = 37$).

Materials

Target sequences in Experiment 2 consisted of Monosyllabic and Disyllabic context versions of target sequences from Experiment 1, except the initial four syllables were removed by splicing at the onset of the 5th syllable of each target sequence. The initial syllable of Experiment 2 target sequences thus corresponded to what had been the 5th syllable of Experiment 1, yielding three types of truncated prosody (Truncated F0, Truncated Duration, and Truncated F0+Duration). As in Experiment 1, the acoustics of the final three syllables were fixed for the monosyllabic and disyllabic versions of all target sequences.

To create filler items for Truncated F0 and Truncated F0+Duration conditions, 32 of 40 filler items from the corresponding conditions from Experiment 1 were truncated at word boundaries by removing 1–6 syllables from the initial part of each stimulus; the remaining eight filler sequences were left intact. To create fillers for the Truncated Duration condition, the same 32 of 40 filler items from the Duration condition from Experiment 1 were subjected to the same splices as for the Truncated F0(+Duration) conditions, while the remaining eight filler items from the Duration condition of Experiment 1 were left intact. Fillers thus ranged in length from 3 to 7 syllables and consisted of two to four real words and no part-words. Splice points were determined by visual inspection of the spectrogram and waveform using Praat; all splices were made at zero-crossings, and the results were checked for naturalness by the first author.

Procedure

The procedure was identical to Experiment 1. Participants listened to 6 practice sequences, followed by 20 targets and 80 fillers—presented in a random order—and freely wrote down the final word that they heard. Half of targets for a participant were in a Monosyllabic context, while the other half were in a Disyllabic context. The target sequence–context pairing was counterbalanced across participants, so that across all participants, each target sequence was tested in both a Monosyllabic and a Disyllabic context.

Results

Fig. 2B shows mean disyllabic response proportions with 95% confidence intervals for Disyllabic and Monosyllabic contexts for Truncated F0, Truncated Duration, and Truncated F0+Duration conditions. Summary results for the associated d' and c measures are shown in Figs. 3A and 3B (white bars), respectively. Inconsistent with the view that the parsing differences observed for Monosyllabic vs. Disyllabic contexts in Experiment 1 were caused solely by 'near distal' prosodic factors, effects of distal prosody were much weaker in Experiment 2 than in Experiment 1. Truncating the syllable sequences examined in Experiment 1 in the F0, Duration and F0+Duration conditions reduced the impact of Disyllabic versus Monosyllabic context on final word reports, as evidenced by smaller mean differences in disyllabic response proportions (Experiment 1: $M = 0.38$, 95% CI = 0.33–0.42; Experiment 2: $M = 0.23$, 95% CI = 0.18–0.28) and smaller d' values (Experiment 1: $M = 1.16$, 95% CI = 1.01–1.31; Experiment 2: $M = 0.69$, 95% CI = 0.53–0.85).

Combining data from Experiments 1 and 2, a 3 (type of prosody: F0, Duration, F0+Duration) × 2 (type of context: Monosyllabic vs. Disyllabic) × 2 (amount of distal context: Experiment 1 vs. 2) mixed measures ANOVA on disyllabic response proportions revealed main effects of type of context ($F_1(1,241) = 345.16$, $MSE = 0.032$, $p < .001$; $F_2(1,19) = 127.38$, $MSE = 0.044$, $p < .001$; $\min F'(1,35) = 93.04$, $p < .001$), type of prosody ($F_1(2,241) = 8.00$, $MSE = 0.091$, $p < .01$; $F_2(2,38) = 17.47$, $MSE = 0.022$, $p < .001$; min-

$F'(2,220) = 5.49$, $p < .01$) and critically, amount of distal context ($F_1(1,241) = 14.93$, $MSE = 0.091$, $p < .001$; $F_2(1,19) = 37.60$, $MSE = 0.017$, $p < .001$; $\min F'(1,157) = 10.69$, $p < .001$). Moreover, there were significant interactions between type of prosody and type of context ($F_1(1,241) = 21.75$, $MSE = 0.032$, $p < .001$; $F_2(2,38) = 57.13$, $MSE = 0.007$, $p < .001$; $\min F'(1,239) = 15.75$, $p < .001$) and between amount of distal context and type of context ($F_1(1,241) = 19.50$, $MSE = 0.032$, $p < .001$; $F_2(1,19) = 34.31$, $MSE = 0.010$, $p < .001$; $\min F'(1,116) = 12.43$, $p < .001$). The latter interaction suggested that truncating the syllable sequences in Experiment 2 (reducing the amount of distal context) had a larger effect on disyllabic response proportions for the Disyllabic context than for the Monosyllabic context. There was additionally an interaction between amount of distal context and type of prosody that was significant by items only ($F_1(2,241) = 1.38$, $MSE = 0.091$, $p = .26$; $F_2(1,19) = 34.31$, $MSE = 0.010$, $p < .01$; $\min F'(2,256) = 1.33$, $p = .27$).

An ANOVA on d' combining the data from the two experiments supported the conclusion that Experiment 2 showed the same pattern of performance across prosody conditions as in Experiment 1, but d' values were significantly lower for all three truncated prosody conditions (see Fig. 3A). This analysis revealed a main effect of type of prosody ($F_1(2,241) = 24.78$, $MSE = 0.629$, $p < .001$; $F_2(2,38) = 19.60$, $MSE = 0.293$, $p < .001$; $\min F'(2,111) = 10.94$, $p < .001$), and a main effect of amount of distal context ($F_1(1,241) = 20.77$, $MSE = 0.629$, $p < .001$; $F_2(1,19) = 28.36$, $MSE = 0.204$, $p < .001$; $\min F'(1,93) = 11.99$, $p < .001$), but no interaction between amount of distal context and type of prosody ($F_1(2,241) = 0.842$, $MSE = 0.629$, $p = .432$; $F_2(2,38) = 0.371$, $MSE = 0.129$, $p = .692$; $\min F'(2,77) = 0.26$, $p = .77$).

Finally, amount of distal context affected response bias. In general, mean values of c for the Truncated F0, Truncated Duration, and Truncated F0+Duration conditions in Experiment 2 were higher than those observed for the associated un-truncated prosody conditions in Experiment 1, indicating a reduced overall tendency to make disyllabic final word reports (see Fig. 3B). This was supported statistically by a 3 (type of prosody) $\times$ 2 (amount of distal context) mixed-measures ANOVA on c, which revealed a main effect of amount of distal context ($F_1(1,241) = 15.79$, $MSE = 0.455$, $p < .001$; $F_2(1,19) = 7.86$, $MSE = 0.366$, $p < .05$; $\min F'(1,42) = 5.25$, $p < .05$), and a (marginally) significant effect of type of prosody, ($F_1(2,241) = 7.29$, $MSE = 0.455$, $p < .001$; $F_2(2,38) = 3.34$, $MSE = 0.569$, $p < .05$; $\min F'(2,78) = 2.29$, $p = .10$), but no interaction between type of prosody and amount of distal context ($F_1(2,241) = 1.62$, $MSE = 0.455$, $p = .20$; $F_2(2,38) = 0.957$, $MSE = 0.501$, $p = .393$; $\min F'(2,91) = 0.60$, $p = .55$).

Discussion

Experiment 2 showed that presenting participants with versions of stimuli from Experiment 1 in which the amount of distal context had been reduced critically weakened effects of distal prosody on word segmentation. Truncating distal context affected proportions of disyllabic final word reports and reduced sensitivity to distal prosody, as evidenced by smaller differences in disyllabic response proportions across Disyllabic and Monosyllabic contexts and lower d' scores. These findings mean that results from Experiment 1 can not be due solely to manipulations at the locus of the 'near distal' 5th syllable, since according to both 'near distal' accounts, there should have been no effect of removing the initial four syllables. Taken together, Experiments 1 and 2 demonstrate a effect of 'far distal' context on word segmentation that is consistent with a perceptual grouping account, whereby both 'near' and 'far' distal prosodic contexts influence the relative strengths of prosodic boundaries around proximal syllables through the generation of expectations that are strengthened with repetition.

One issue raised by Experiments 1 and 2 is that both studies involved an explicit task in which participants were asked to write down the final word they heard in each syllable sequence. As such, it is not clear whether effects of distal prosody might also be observed in a task which does not require such explicit, and possibly metalinguistic, processing. To address this issue, a third experiment was conducted using a study-test recognition design that did not involve an explicit word segmentation judgment.

In Experiment 3, participants first performed a phoneme-monitoring task for a set of target syllable sequences presented in Monosyllabic and Disyllabic contexts plus an additional set of filler sequences; they were then given a surprise recognition test involving visually-presented disyllabic items. Based on the perceptual grouping account, we expected that the strength of encoding of a lexical item and accuracy of later recognition should be modulated by distal prosody in a manner consistent with the implied perceptual grouping and prosodic constituency. Specifically, individuals should better recognize *visually*-presented disyllabic words composed of adjacent syllables in previously *heard* sequences (e.g., *notebook* in *note-book-worm*) when the distal prosodic context was predicted to facilitate perceptual grouping of those syllables into a single disyllabic word, than when it was not predicted to facilitate such a grouping. An additional goal of Experiment 3 was to test the prediction of the perceptual grouping hypothesis that distal prosodic context should affect not only the perceived lexical and prosodic constituency of the final 7th and 8th syllables, but also that of syllables earlier in the sequence (e.g., syllables 5, 6, and 7).

Experiment 3

Methods

Participants

Seventy-eight native speakers of American English completed the experiment in return for course credit in an introductory psychology course at the Ohio State University. Participants were at least 18 years of age with self-reported normal hearing and a range a musical experience.

Design

There were two parts to the experiment. The first part (the study phase) consisted of an offline phoneme monitoring task involving target and filler syllable sequences. The second part (the test phase) consisted of a 'surprise' word recognition test with *visually*-presented disyllabic items. For the latter, there was a single within-subject factor, item type, which had four levels: Congruent, Incongruent, Neutral, and New. Congruent items were visually-presented disyllabic words that were heard in the study phase as adjacent syllables in a prosodic context predicted to facilitate perceptual grouping of those syllables into a single disyllabic word. Incongruent items were disyllabic words that were also heard as adjacent syllables in the study phase, but in a prosodic context predicted to inhibit their perceptual grouping. For example, for the sequence ending with *foot-note-book-worm*, *footnote* was a Congruent item when presented in a Disyllabic context (since the parsing was expected to be *footnote bookworm*, with *bookworm* being the predicted response in Experiment 1), while *notebook* was an Incongruent item in this context. Conversely, *footnote* was an Incongruent item when presented in a Monosyllabic context (since the parsing was expected to be *foot notebook worm*, with *worm* being the predicted response in Experiment 1), while *notebook* was a Congruent item in this context. Each Congruent or Incongruent item always corresponded to the disyllabic word formed by syllables 5–6 or 6–7. Neutral items were disyllabic words that were heard in the study phase as adjacent syllables with unambiguous lexical organization. New items were disyllabic words that were not heard during the study phase.

Materials

Stimuli for the phoneme monitoring task (study phase) consisted of monosyllabic and disyllabic versions of the 10 eight-syllable target sequences that generated the largest d' scores in a by-items analysis of the F0+Duration condition of Experiment 1 (the first 10 target sequences in the Appendix A). Moreover, there were 40 filler sequences of varying lengths, also selected from the F0+Duration condition of Experiment 1.

Stimuli for the recognition test consisted of 40 visually-presented disyllabic words. Twenty of these words constituted Congruent and Incongruent items, which were created by concatenating adjacent syllables 5 and 6 (e.g., to form *footnote*) and 6 and 7 (e.g., to form *notebook*) in the target sequences. Neutral items were disyllabic words selected from the filler sequences. Nine of 10 Neutral items occurred in second or third to last position from the end of the filler sequences, while one item was fourth from the end. Finally, New items were disyllabic, compound words which had not occurred in target or filler sequences during the study phase.

Procedure

For the phoneme monitoring task (study phase), participants were instructed they would hear lists of words, and that for each word list they should circle 'yes' or 'no' to indicate whether the sequence contained a /b/ sound, as at the beginning of *big* or the end of *lab*. Participants listened to six filler sequences as practice, followed by 10 target sequences and 40 filler sequences presented in random order. Half of the target sequences were paired with a Monosyllabic context, while the other half were paired with a Disyllabic context. Target sequence-context pairing was counterbalanced across participants. Two lists were constructed; each target sequence occurred in one prosodic context in one list, and the other prosodic context in the other list. Two more lists were created by reversing the order of these lists, for a total of four lists. Each participant was randomly assigned to one of the four lists, with approximately equal numbers in each list.

For the word recognition task (test phase), participants were instructed that they would see a word presented on the computer screen which may or may not have occurred in the preceding word list they heard. They circled 'yes' on a paper answer sheet if they heard the word earlier in the wordlist during the phoneme monitoring task and 'no' if they did not hear the word earlier. The recognition test consisted of 40 disyllabic words (10 Congruent items, 10 Incongruent items, 10 Neutral items, and 10 New items); the same lexical item (e.g., *footnote*) was previously heard as either a Congruent or an Incongruent stimulus, depending on the prosodic context in which it was experienced in the study phase. Words were presented visually on a computer screen with an inter-stimulus-interval of 4 s. The 40 disyllabic words were presented in one random order to half the participants and in the reverse order to the remaining participants. The experiment lasted approximately 30 min.

Results and discussion

Because the acoustic characteristics of syllable 5, but not syllable 6, varied as a function of prosodic context, responses were initially separated by position of the syllable pair in the target sequence (5–6 vs. 6–7), in order to assess whether there was an impact of syllable position on recognition performance. No effect of syllable-pair position was found in any of the analyses. Consistent with the perceptual grouping hypothesis, participants better recognized Congruent disyllabic items (syllable pair 5–6: $M = 0.60$, 95% CI = 0.49–0.71; syllable pair 6–7; $M = 0.63$, 95% CI = 0.47–0.79) than Incongruent disyllabic items (syllable pair 5-6: $M = 0.37$, 95% CI = 0.21–0.54; syllable pair 6–7: $M = 0.39$, 95% CI = 0.23–0.54) for both syllable-pair positions. The recognition difference between Congruent and Incongruent items was supported by a 2 (item type: Congruent versus Incongruent) $\times$ 2 (position: syllable pair 5–6 versus syllable pair 6–7) repeated measures ANOVA, which revealed a main effect of item type ($F_1(1,77) = 58.06$, $MSE = 0.072$, $p < .001$; $F_2(1,9) = 331.23$, $MSE = 0.002$, $p < .001$; $\min F'(1,84) = 49.4$, $p < .001$), but no effect of position of the disyllabic item ($F_1(1,77) = 0.716$, $MSE = 0.040$, $p = .4$; $F_2(1,9) = 0.06$, $MSE = 0.073$, $p = .82$; $\min F'(1,11) = 0.05$, $p = .82$) or significant interaction between item type and position ($F_1(1,77) = 0.253$,

$MSE = 0.041$, $p = .62$; $F_2(1,9) = 0.04$, $MSE = 0.022$, $p = .85$; $\min F'(1,12) = 0.04$, $p = .86$).

Next, a signal detection analysis was undertaken; for this analysis, correct recognition responses (i.e., 'yes' responses to disyllabic test items that were previously heard during the phoneme monitoring task) were coded as hits, while false recognition responses (i.e., 'yes' responses to New disyllabic items not previously heard during the phoneme monitoring task) were coded as false alarms. Based on hits and false alarms we calculated d' and c to index (1) participants' ability to discriminate in memory between previously heard and New disyllabic items and (2) any tendency participants showed to simply respond 'yes' on the recognition test, respectively. Mean d' and c with 95% confidence intervals for Congruent, Incongruent, and Neutral disyllabic test items are shown in Fig. 4 (A and B, respectively). Separate one-way ANOVAs on d' and c both showed main effects of item type ($F_1(2,154) = 37.93$, $p < .001$ for the analysis of d', and $F_1(2,154) = 37.28$, $MSE = 0.007$, $p < .001$ for the analysis of c).

Critically, test items that had been previously heard paired with congruent contexts (Congruent items) yielded the highest values of d' (best recognition), Incongruent items produced the lowest values of d' (worst recognition), while Neutral items (previously heard as fillers) produced intermediate values of d' (see Fig. 4A). Paired-samples t-tests revealed significant differences in recognition memory for all pairs of conditions (Congruent versus Incongru-

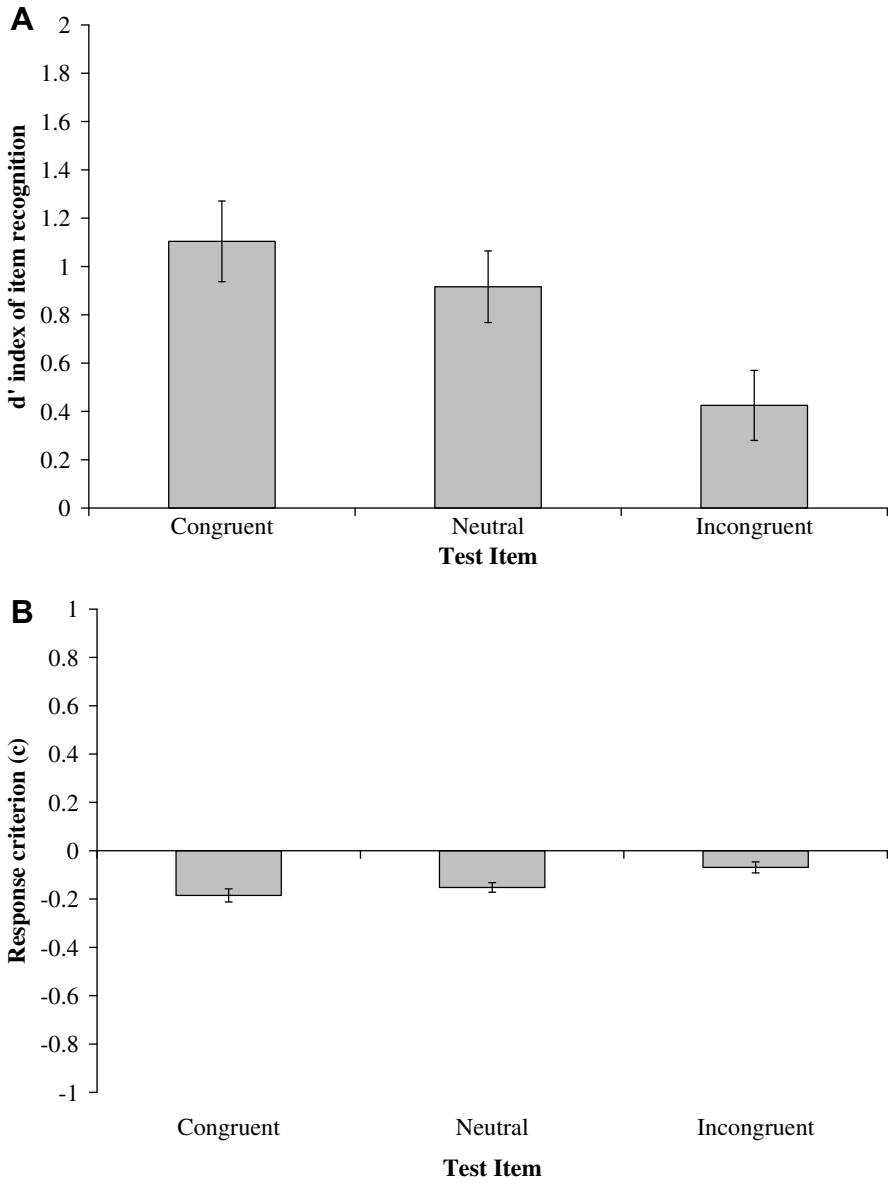

Fig. 4. Signal detection measures of recognition performance in Experiment 3. (A) Mean values of d' with 95% confidence intervals for Congruent, Neutral and Incongruent test items. (B) Mean values of c with 95% confidence intervals for Congruent, Neutral and Incongruent test items.

ent: $M = 0.68$, 95% CI = 0.50–0.86, $t_1(77) = 7.65$, $p < .01$; Congruent versus Neutral: $M = 0.19$, 95% CI = 0.06–0.32, $t_1(77) = 2.931$, $p < .01$; Incongruent versus Neutral: $M = -0.49$, 95% CI = −0.66 to −0.32, $t_1(77) = -5.69$, $p < .01$). Overall, mean values of *c* were slightly negative for all three item types, indicating a small, but significant, bias to respond 'yes' to test items independent of whether they had been previously heard or not (see Fig. 4B).

In sum, results of Experiment 3 show that participants' ability to recognize previously heard lexical items depended on distal prosodic context. Participants better recognized that they previously heard a disyllabic word when distal prosody was predicted to facilitate grouping those syllables (i.e., for Congruent items) than when distal prosody was not predicted to facilitate grouping those syllables (i.e., for Incongruent items). Moreover, distal prosody was found not only to affect the final two syllables, but also syllables occurring earlier in the sequence. Overall, findings are consistent with the view that distal prosodic context affected perceptual grouping and proximal prosodic constituency, and therefore word segmentation, where this took place at an unconscious, implicit level of processing. Because participants did not know they were going to be given a recognition test, these data suggest that it is unlikely that the distal prosodic effects demonstrated in Experiments 1 and 2 arose from a conscious or metalinguistic processing strategy. In summary, Experiment 3 has provided converging evidence for the perceptual grouping hypothesis using a complementary study-test recognition paradigm.

General discussion

Three experiments identify distal prosody as a new factor in word segmentation and lexical processing and provide broad support for a perceptual grouping hypothesis; this hypothesis proposed that distal prosodic context affected the relative strengths of proximal prosodic phrase boundaries in accordance with general principles of auditory perceptual organization. In all three experiments, participants listened to ambiguous target syllable sequences (e.g., *footnote bookworm*, *foot notebook worm*), while distal F0 and/or duration patterns of speech were manipulated. In both Experiments 1 and 2, participants listened to target and filler sequences and reported the last word they heard in each sequence. In Experiment 3, participants completed a phoneme monitoring task involving the target syllable sequences and were then given a 'surprise' word recognition test that contrasted recognition performance for Congruent and Incongruent items (i.e., disyllabic items that were previously heard in either a congruent or incongruent distal prosodic context).

There are five main findings from these experiments. First, identical acoustic material was perceived as different lexical items, depending on distal F0 and/or duration cues (Experiments 1 and 2). Disyllabic distal prosodic contexts led to disyllabic final word reports, whereas Monosyllabic distal prosodic contexts led to monosyllabic final word reports. Second, combined F0 and duration cues yielded stronger effects than either F0 or duration cues alone (Experiments 1 and 2). Third, reducing the amount of distal prosodic context weakened the effect of distal prosody on participants' final word reports (Experiment 2). Fourth, distal prosody affected participants' later recognition of target items, even though identical acoustic material was presented (Experiment 3); participants better recognized visually-presented disyllabic words when the distal prosodic context heard earlier was predicted to facilitate grouping of syllables into a disyllabic word, than when the distal context was predicted not to facilitate this grouping. Finally, the effect of distal prosody on to-be-recognized candidate disyllabic items extended to syllable pairs other than those ending each target sequence (Experiment 3).

In addition to providing support for the perceptual grouping hypothesis, the reported experiments provide evidence against several alternative explanations. First, the present results are inconsistent with an explanation that the observed distal segmentation effects are due to a dispreference for parsings that generate a proximal 'low pitch accent' on a main stress syllable. In prosodic conditions that manipulated F0, a lexical parsing yielding a disyllabic final word entailed that the main stress of that word (e.g., *book* in *bookworm*) was always low in F0 (i.e., it was a low pitch accent). Thus, if individuals had a dispreference for parsings generating low pitch accents, they should have disproportionately given monosyllabic final word reports, regardless of distal context. This was not the case, as signal detection analyses revealed that the mean estimated response criterion, *c*, was approximately zero for both the F0 and the F0+Duration conditions (i.e., monosyllabic and disyllabic final word reports occurred in approximately equal numbers). Moreover, an account based on a dispreference for low pitch accents would necessarily provide only a partial explanation of the results, since this would not explain distal segmentation effects in conditions where pitch was held constant (i.e., the Duration conditions in Experiments 1 and 2); in these conditions, the final three syllables were monotone and perceptually isochronous.

The present experiments also provide evidence against variants of 'near distal' accounts that propose that acoustic cues on the 'near distal' 5th syllable either alone or in combination are responsible for the observed prosody-based segmentation effects. The F0 change and/or durational lengthening that occurred on the 5th syllable likely induced perception of a major prosodic phrase boundary after that syllable (cf. Shattuck-Hufnagel & Turk, 1996; Turk & Sawusch, 1997; Turk & Shattuck-Hufnagel, 2000; Turk & White, 1999). For example, the view that either F0 or duration cues *alone* on the 'near distal' 5th syllable were responsible for context-dependent segmentation patterns is untenable for the straightforward reason that Experiments 1 and 2 showed that combined distal F0 and duration cues resulted in greater parsing differences than either cue alone. More generally, all 'near distal' explanations involving F0 alone, duration alone, or both cues combined incorrectly predict that removing the first four syllables of context should have no effect on the strength of distal prosodic effects on segmentation. In Experiment 2, we found that reducing the amount of distal prosody

by removing the first four syllables of context reduced the magnitude of the effect compared to Experiment 1 for all three distal prosody conditions.

One possible way to salvage a 'near distal' account is to propose that enhanced effects of 'far distal' context observed for Experiment 1 relative to Experiment 2 are due to an improved ability with greater context to attribute lengthening on the 5th syllable to a major prosodic boundary, e.g., a full intonation phrase boundary (Shattuck-Hufnagel & Turk, 1996). According to this explanation, lengthening e.g., on *foot* might have been more readily identified as a major prosodic boundary with four preceding syllables (Experiment 1) than with zero preceding syllables (Experiment 2). Better identification of large prosodic boundaries could lead to more consistent application of a strategy such as grouping subsequent syllables according to the longest initial lexical candidate, thereby accounting for enhanced effects of context on segmentation in Experiment 1. However, this explanation is at best only a partial explanation, because a weakened effect on segmentation was critically observed in the F0 condition as well, which contained no duration manipulation. Moreover, it is not clear why such 'far distal' effects would be predicted under existing prosodic accounts, which are not based on perceptual grouping.

Finally, Experiment 3 provided converging evidence for the perceptual grouping hypothesis by using a complementary task involving a study-test recognition design. In this experiment, better recognition of target disyllabic words in a surprise memory test was found for items heard earlier paired with congruent distal contexts than for identical target items heard earlier paired with incongruent distal contexts. Because participants were unaware that they would face a recognition test, Experiment 3 suggests that distal prosody operates at an unconscious, implicit level.

In the remainder of this article, we situate the present set of findings supporting a perceptual grouping account of effects of distal prosody on word segmentation and lexical processing in the context of previous research and prosodic theory.

Implications for lexical processing and lexical access

Previous studies of prosodic effects on lexical processing have almost exclusively investigated proximal prosodic characteristics (e.g., Cutler & Donselaar, 2001; Soto-Faraco et al., 2001; van Donselaar et al., 2005; Davis et al., 2002; Donsela et al., 2005; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). The present study adds to this work by demonstrating clear effects of *distal* prosody on lexical processing. In the present study, distal prosodic cues influenced both explicit lexical parsing judgments, as well as lexical processing as gauged by recognition memory performance. These experiments clearly show that both distal F0 and distal durational cues separately affected lexical organization, as indicated in Experiments 1 and 2. The stronger effect observed for distal F0 cues relative to distal durational cues in both experiments could be due to a perceptual dominance of frequency over duration cues. Perception of metrical stress, which is a lexically relevant property, has been shown to depend more on frequency than on duration (e.g., Fry, 1958); consistent with this, proximal stress-based segmentation is affected more by neutralizing F0 information than neutralizing duration information (Spitzer et al., 2007).[3] Similarly, fundamental frequency cues appear to be more important than temporal cues in conveying metrical structure in music (Hannon et al., 2004). These earlier findings are consistent with the result that F0 provided a stronger cue to lexical organization than duration in the present experiments.

The observed distal prosodic effects also challenge the widespread assumption that prosodic cues are minimally useful in lexical access, since they are necessarily available to the perceptual system later than segmental or coarticulatory information; see Cooper, Cutler, and Wales (2002) for a discussion of this view. This position has been maintained on the basis that proximal segmental and coarticulatory information is utilized by the listener almost as rapidly as the speaker produces it (Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999; Strange, 1989; Whalen, 1991). In contrast, proximal prosodic cues, e.g., to stress, are highly variable (Cutler & Norris, 1988; Mattys, 2004; Sluijter & van Heuven, 1996). The instability of proximal prosodic cues, together with the ability to rapidly perceive segmental/coarticulatory ones, has led to the view that the latter cues provide the earliest useful information for lexical access. However, the present work suggests that distal prosodic information *several syllables in advance of* proximal segmental and coarticulatory cues can potentially influence lexical processing, so that prosodic information may be more useful and/or influential in processing than previously believed.

Finally, although a number of studies of lexical processing have used compound words or phrases (e.g., Banel & Bacri, 1994; Gow & Gordon, 1995), we speculate that even stronger effects of distal prosody might have been observed if non-compound disyllabic words with end-embedding had been used (e.g., *surplus* vs. *plus*). The relatively obvious end-embedding of final monosyllabic words in disyllabic compounds may have increased the rate of monosyllabic reports in the Disyllabic context, thereby weakening an already strong effect. Additional research is needed to address this issue.

Implications for word segmentation

A second area of research in which effects of proximal prosody have been of interest is word segmentation. Segmentation based on metrical prosody has been documented in a large number of studies (see Cutler, Dahan, & Donselaar, 1997 for a review). In particular, stressed syllables tend to be perceived as word-initial in English and other stressed-timed languages (Cutler & Butterfield 1992; Cutler & Norris, 1988; Norris, McQueen, & Cutler, 1995), while proximal durational cues also play a role in

[3] It is interesting to note that in the Spitzer et al. (2007) study, neutralizing both proximal F0 and proximal durational cues together did not result in less reliance on stress-based segmentation relative to neutralizing either proximal cue alone. This contrasts with the present study, in which the combination of both distal F0 and distal durational cues resulted in additive segmentation effects relative to either distal cue alone.

lexical parsing (Banel & Bacri, 1994; Nakatani & Schaffer, 1978). The present results suggest that distal prosody upstream of the locus of segmentation also affects word segmentation, indicating a number of new avenues for future research.

In addition to signal-based cues, such as proximal prosody, it is well-known that knowledge-based cues, including semantics, syntax, etc., play an important role in word segmentation. Recently, the relative strengths of different types of word segmentation cues have been investigated by Mattys and colleagues (Mattys, 2004; Mattys & Melhorn, 2007; Mattys, White, & Melhorn, 2005; Mattys et al., 2007). This research has shown that, for adult listeners at least, (proximal) prosody is outranked by most other types of cues, including allophonic variation, semantics, lexicality, etc., except under conditions of signal degradation (i.e., noise).

Based on these results, Mattys et al. (2005) proposed a hierarchical framework for word segmentation, which provided a ranking of proximal prosody (i.e., stress cues) relative to other segmentation cues. While this framework accounts for many facets of listener segmentation behavior, it does not account for the present results, since no role for distal prosodic cues is posited. Additional work will therefore be needed to determine the ranking of distal prosodic cues relative to other word segmentation cues.

It should be noted that while knowledge-based cues are quite important for adult listeners, they are much less available for infants. The significance of proximal prosodic characteristics in infant word segmentation is well-attested (see Saffran, Werker, & Werner, 2006, for a review). We propose that distal prosodic cues potentially play an important role in infant word segmentation by creating prosodic phrasings that favor particular word candidates for proximal syllables. Moreover, Experiment 3 suggests that distal prosodic context may affect word learning. Such possibilities may be fruitfully examined in future studies.

Moreover, the present results suggest a cautionary note in interpreting prior studies investigating effects of proximal prosodic differences on lexical access and word segmentation. In particular, several of these studies used different recordings and/or word strings for sentential contexts of critical paired target words and thus failed to control for distal prosodic cues (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001; Cooper et al., 2002; van Donselaar et al., 2005). Our results suggest that distal prosody might partially account for findings attributed to proximal prosody (i.e., to stress differences). A number of earlier studies (e.g., Bond & Small, 1983; Cutler, 1986; Cutler & Clifton, 1984; Slowiaczek, 1990, 1991; Small, Simon, & Goldberg, 1988) had failed to find effects of proximal prosody on processing; this would seem to increase the likelihood that distal characteristics may have played a role in more recent studies attributing effects to proximal prosody. The present findings therefore indicate that distal prosody will need to be controlled in future studies examining proximal prosody.

Implications for theories of speech prosody

In this section we consider how these findings fit with the theory of the prosodic hierarchy (Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986). In order of increasing constituent size, prosodic constituent types that have been proposed to be part of this hierarchy include (but are not limited to) the syllable, the foot, the prosodic word, the phonological phrase, the intermediate intonation phrase, and the full intonation phrase (Beckman & Pierrehumbert, 1986; Hayes & Lahiri, 1991), although there is some disagreement regarding which levels of the hierarchy are distinguished (Shattuck-Hufnagel & Turk, 1996). Our results indicate that distal prosodic cues affected perceived proximal prosodic constituency, since distal prosodic context affected perceived locations of proximal lexical boundaries, which correspond to prosodic boundaries at the level of the prosodic word or higher constituent (Shattuck-Hufnagel & Turk, 1996).

Our results also provide evidence against a frequent assumption in linguistics and psychology, namely that prosodic constituency is cued solely by proximal acoustic characteristics. On the one hand, there is plenty of evidence that larger prosodic constituents tend to be cued by larger proximal acoustic changes, e.g., increased local durational lengthening, a larger F0 drop, and/or increased use of glottalized voicing (Fougeron & Keating, 1997; Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Cho & McQueen, 2005; Turk & Shattuck-Hufnagel, 2000). However, our results indicate that distal prosodic characteristics also contribute to perceived prosodic constituency, even when proximal acoustic cues are held constant.

What sorts of prosodic boundaries might listeners have been hearing proximally? Lexical word boundaries correspond to prosodic boundaries at the level of the prosodic word or higher constituent (Shattuck-Hufnagel & Turk, 1996). We propose that distal prosodic context aids the interpretation of prosodic structure of syllables according to a parallelism principle (Dilley, 2008; Dilley & McAuley, 2006; Dilley & Shattuck-Hufnagel, 1999). According to this principle, sequences of syllables preferably form parallel parts of groups with parallel metrical structure, building on a similar proposal for music (Lerdahl & Jackendoff, 1983). In English, accents can be high or low, and phrasal boundaries can also be high or low (e.g., Pierrehumbert, 1980). The stimuli used in the present experiments exploited this property of English by alternating sequences of syllables with high and low pitch which were acoustically compatible proximally with more than one possible prosodic organization. Lexical items in lists are held to be demarcated by quite large prosodic boundaries – either intermediate or full intonation phrase boundaries – depending on the proposal (Beckman & Ayers Elam, 1997; Liberman & Pierrehumbert, 1984; Shattuck-Hufnagel & Turk, 1996). Regardless of the type of large prosodic boundary that was assumed to demarcate distal lexical items in our materials, the parallelism proposal suggests that listeners interpreted the proximal prosody in a manner which was parallel to the distal prosody, hearing proximal lexical items as demarcated by the same large constituent boundary as occurred distally. In other words, by this account whether or not a given syllable boundary was heard as the location of a large prosodic boundary or not depended solely on the distal prosodic context. Finally, when proximal syllable boundaries were heard as locations of relatively smaller prosodic constituent boundaries

(e.g., the boundary before *worm* when heard as part of *bookworm*), those boundaries were likely to have been foot or smaller boundaries since monosyllabic content words like *book* and *worm* are held to be stressed syllables, and thus are monosyllabic feet in their own right (Shattuck-Hufnagel & Turk, 1996).[4]

Conclusions

In summary, three experiments demonstrated that the perceived lexical organization of identical acoustic material in lexically ambiguous syllable sequences depends on distal prosodic factors. Support was found for a perceptual grouping hypothesis in which distal prosodic characteristics establish perceived patterns of pitch and rhythm that affect the relative strengths of prosodic boundaries at edges of proximal syllables, thereby affecting both word segmentation and lexical processing. Though stimuli with such extensive lexical ambiguity are unlikely to occur often in everyday speech, sequences with ambiguous structure reveal the nature of processes in spoken language understanding which would normally be obscured by other factors. These processes are likely to be important when parsing spoken language in noisy environments, in the early stages of linguistic development, and/or in word learning associated with second language acquisition.

Appendix A

channel dizzy foot *note book worm*
worthy vinyl life *long hand shake*
victim fragile sun *spot light house*
wallet ruthless back *space suit case*
skirmish princess side *kick stand still*
prelude charcoal touch *down play boy*
climate humble wise *crack pot hole*
quaker trophy step *son/Sun day break*
cherry vapor car *pool side walk*
pennies handy half *back fire wood*
rumor habit dream *boat yard stick*
texture nozzle coat *tail gate post*
swivel witness thread *bare foot bridge*
shower cannon pass *word play pen*
genius planet work *horse whip lash*
splendor radish friend *ship board room*
brandy curtain ear *drum head piece*
border taxi life *boat house top*
chocolate lyric down *town ship wreck*
sonar baggy air *field work day*

[4] H. Giegerich (personal communication to the first author) has proposed that the word-medial foot boundary in a compound lexical item like *bookworm* heard proximally might be subject to a process of 'defooting,' whereby this foot boundary is demoted to a syllable boundary in the present distal contexts. Since most distal words in our experimental materials (e.g., *channel*) contained word-medial syllable boundaries which were not themselves foot boundaries, such a 'defooting' process would grant 'parallel' status to proximal word-medial prosodic boundaries compared with distal word-medial prosodic boundaries.

References

Banel, M. H., & Bacri, N. (1994). On metrical patterns and lexical parsing in French. *Speech Communication, 15*, 115–126.

Beckman, M., & Ayers Elam, G. (1997). *Guidelines for ToBI labeling* (Ver. 3.0). The Ohio State University. Available at: <http://www.ling.ohio-state.edu/~tobi/ame_tobi/annotation_conventions.html>.

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

Boersma, P., Weenink, D. (2002). *Praat, a system for doing phonetics by computer*. Software and manual available online at: <http://www.praat.org>.

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

Bond, Z. S., & Small, L. H. (1983). Voicing, vowel and stress mispronunciations in continuous speech. *Perception & Psychophysics, 34*, 470–474.

Bregman, A. S. (1978). Auditory streaming is cumulative. *Journal of Experimental Psychology: Human Perception and Performance, 4*(3), 380–387.

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

Chafe, W. (1988). Linking intonation units in spoken English. In J. Harman & S. Thompson (Eds.), *Clause combining in grammar and discourse*. Amsterdam/Philadelphia: John Benjamins Publishing Company.

Cho, T., & McQueen, J. M. (2005). Prosodic influences on consonant production in Dutch: Effects of prosodic boundaries, phrasal accent and lexical stress. *Journal of Phonetics, 33*, 121–157.

Cho, T., McQueen, J. M., & 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., 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.

Cole, R. A., & Jakimik, J. (1980). Segmenting speech into words. *Journal of the Acoustical Society of America, 64*, 1323–1332.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: evidence from native and nonnative listeners. *Language and Speech, 45*, 207–228.

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

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

Crystal, D., & Quirk, R. (1964). *Systems of prosodic and paralinguistic features in English. Janua Linguarum*. London, The Hague: Mouton.

Cutler, A. (1976). Phoneme-monitoring reaction time as a function of preceding intonation contour. *Perception & Psychophysics, 20*, 55–60.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*, 201–220.

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

Cutler, A., & Clifton, C. E. (1984). The use of prosodic information in word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and performance X* (pp. 183–196). Hillsdale, NJ: Erlbaum.

Cutler, A., Dahan, D., & Donselaar, W. van. (1997). Prosody in the comprehension of spoken language: A literature review. *Language and Speech, 40*, 141–202.

Cutler, A., & van Donselaar, W. (2001). Voornaam is not a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*, 171–195.

Cutler, A., & Norris, D. G. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121.

Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218–244.

Dilley, L. (1997). Some factors influencing duration between syllables judged perceptually isochronous. *Journal of the Acoustical Society of America, 102* (5), Pt.2, 3205–3206.

Dilley, L. (2008). Empirical perspectives on prosodic structure: A theoretical appraisal. *Presentation at Experimental and theoretical Advances in Prosody, Cornell University*. April 12, pp. 3205–3206.

Dilley, L. & McAuley, J. D. (2006). Perceptual organization in intonational phonology: A test of parallelism. Presented at the *10th laboratory phonology conference*, Paris, France.

Dilley, L. & Shattuck-Hufnagel, S. (1999). Effects of repeated intonation patterns on perceived word-level organization. In *Proceedings of the 14th international congress of phonetic sciences*, San Francisco, Vol. I, pp. 1487–1490.

Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of vowel-initial syllables as a function of prosodic structure. *Journal of Phonetics, 24*, 423–444.

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

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *Quarterly Journal of Experimental Psychology, 58A*, 251–273.

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

Fry, D. B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*, 765–768.

Fry, D. B. (1958). Experiments in the perception of stress. *Language and Speech, 1*, 126–152.

Gibbon, D. (1976). *Perspectives of intonational analysis*. Bern: Herbert Lang.

Gómez, R. (2002). Variability and detection of invariant structure. *Psychological Science, 13*(5), 431–436.

Gómez, R., & Maye, J. (2005). The developmental trajectory of nonadjacent dependency learning. *Infancy, 7*(2), 183–206.

Gout, A., Christophe, A., & Morgan, J. (2004). Phonological phrase boundaries constrain lexical access. II: Infant data. *Journal of Memory and Language, 51*, 548–567.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 344–359.

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

Halliday, M. A. K. (1967). *Intonation and grammar in British English*. The Hague: Mouton.

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

Hannon, E. E., Snyder, J. S., Eerola, T., & Krumhansl, C. L. (2004). The role of melodic and temporal cues in perceiving musical meter. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 956–974.

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

Jones, M. R. (1976). Time, our lost dimension: Toward a new theory of perception, attention, and memory. *Psychological Review, 83*(5), 323–355.

Jones, M. R., & Boltz, M. (1989). Dynamic attending and responses to time. *Psychological Review, 96*(3), 459–491.

Jones, M. R., & Yee, W. (1993). Attending to auditory events: The role of temporal organization. In S. McAdams & E. Bigand (Eds.), *Thinking in sound: The cognitive psychology of human audition* (pp. 199–230). Oxford, England: Oxford University Press.

Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, The Ohio State University.

Kidd, G. R. (1989). Articulatory rate–context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*(4), 736–748.

Kingdon, R. (1958). *The groundwork of English intonation*. London: Longman.

Klatt, D. H. (1980). Speech perception: A model of acoustic–phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243–288). Hillsdale, NJ: Erlbaum.

Ladd, D. R. (1986). Intonational phrasing: the case for recursive prosodic structure. *Phonology Yearbook, 3*, 311–340.

Ladd, D. R. (1996). *Intonational phonology*. Cambridge, MA: Cambridge University Press.

Large, E. W., & Jones, M. R. (1999). The dynamics of attending: How people track time-varying events. *Psychological Review, 106*, 119–159.

Large, E. W., & Palmer, C. (2002). Perceiving temporal regularity in music. *Cognitive Science, 26*, 1–37.

Lehiste, I. (1970). *Suprasegmentals*. Cambridge, MA: MIT Press.

Lehiste, I. (1977). Isochrony revisited. *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.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: Words, phonemes, and features. *Psychological Review, 101*, 653–675.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions during word recognition in continuous speech. *Cognition, 10*, 29–63.

Martin, J. G. (1972). Rhythmic (hierarchical) versus serial structure in speech and other behavior. *Psychological Review, 79*, 487–509.

Martin, J. G. (1979). Rhythmic and segmental perception are not independent. *Journal of the Acoustical Society of America, 65*, 1286–1297.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception & Psychophysics, 62*, 253–265.

Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 397–408.

Mattys, S. L., Jusczyk, P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465–494.

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.

Mattys, S. L., Melhorn, J. F., & White, L. (2007). Effects of syntactic expectations on speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 33*, 960–977.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*(4), 477–500.

McAuley, J. D., & Dilley, L. (2004). Acoustic correlates of perceived rhythm in spoken English. *Journal of the Acoustical Society of America, 115*, 2397.

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

McAuley, J. D., & Kidd, G. R. (1998). Effect of deviations from temporal expectation on tempo discrimination of isochronous tone sequences. *Journal of Experimental Psychology: Human Perception and Performance, 24*(6), 1786–1800.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.

McQueen, J. M., Norris, D. G., & Cutler, A. (1999). Lexical influence in phonetic decision-making: Evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance, 25*, 1363–1389.

Meltzer, R. H., Martin, J. G., Mills, C. B., Imhoff, D. L., & Zohar, D. (1976). Reaction time to temporally displaced phoneme targets in continuous speech. *Journal of Experimental Psychology: Human Perception and Performance, 2*, 277–290.

Millotte, S., René, A., Wales, R., & Christophe, A. (2008). Phonological phrase boundaries constrain the online syntactic analysis of spoken sentences. *To appear in Journal of Experimental Psychology: Learning, Memory, and Cognition*.

Morgan, J. L. (1996). A rhythmic bias in preverbal speech segmentation. *Journal of Memory and Language, 35*, 666–688.

Moulines, E., & Charpentier, F. (1990). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication, 9*(5-6), 453–467.

Nakatani, L. H., & Schaffer, J. A. (1978). Hearing 'words' without words: Prosodic cues for word perception. *Journal of the Acoustical Society of America, 63*, 234–245.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.

Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.

Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*(5), 1209–1228.

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191–243.

Palmer, H. (1922). *English intonation, with systematic exercises*. Cambridge: Heffer.

Parncutt, R. (1994). A perceptual model of pulse salience and metrical accent in musical rhythms. *Music Perception, 11*, 409–464.

Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT, Cambridge, MA.

Pierrehumbert, J. (2000). Tonal elements and their alignment. In M. Horne (Ed.), *Prosody: Theory and experiment* (pp. 11–36). Dordrecht: Kluwer Academic Publishers.

Pike, K. (1945). *The intonation of American English*. Michigan, USA: University of Michigan Press.
Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 564–573.
Povel, D. J., & Essens, P. (1985). Perception of temporal patterns. *Music Perception, 2*(4), 411–440.
Prieto, P., van Santen, J., & Hirschberg, J. (1995). Tonal alignment patterns in Spanish. *Journal of Phonetics, 23*, 429–451.
Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331–350.
Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America, 94*, 2027–2035.
Quené, H., & Port, R. F. (2005). Effects of timing regularity and metrical expectancy on spoken-word perception. *Phonetica, 62*(1), 1–13.
Saffran, J. R., Werker, J. F., & Werner, L. A. (2006). The infant's auditory world: Hearing, speech, and the beginnings of language. In R. Siegler & D. Kuhn (Eds.), *Sixth edition of the handbook of child development* (pp. 58–108). New York: Wiley.
Salverda, A. ., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51–89.
Salverda, A. P., Dahan, D., Tanenhaus, M. K., Crosswhite, K., Masharov, M., & McDonough, J. (2007). Effects of prosodically-modulated sub-phonetic variation on lexical competition. *Cognition, 105*(2), 466–476.
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.
Schubiger, M. (1958). *English intonation, its form and function*. Tuebingen: Max Niemeyer Verlag.
Selkirk, E. (1984). *Phonology and syntax*. Cambridge, MA: MIT Press.
Shattuck-Hufnagel, S., & Turk, A. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research, 25*(2), 193–247.
Shatzman, K. B., & McQueen, J. M. (2006). Prosodic knowledge affects the recognition of newly acquired words. *Psychological Science, 17*, 372–377.
Shields, J. L., McHugh, A., & Martin, J. G. (1974). Reaction time to phoneme targets as a function of rhythmic cues in continuous speech. *Journal of Experimental Psychology, 102*, 250–255.
Slowiaczek, L. M. (1990). Effects of lexical stress in auditory word recognition. *Language and Speech, 33*, 47–68.
Slowiaczek, L. M. (1991). Stress and context in auditory word recognition. *Journal of Psycholinguistic Research, 20*, 465–481.
Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100*(4), 2417–2485.
Small, L. H., Simon, S. D., & Goldberg, J. S. (1988). Lexical stress and lexical access: Homographs versus nonhomographs. *Perception & Psychophysics, 44*, 272–280.
Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45*, 412–432.
Spitzer, S. M., Liss, J. M., & Mattys, S. L. (2007). Acoustic cues to lexical segmentation: A study of resynthesized speech. *Journal of the Acoustical Society of America, 122*(6), 3678–3687.
Strange, W. (1989). Dynamic specification of coarticulated vowels spoken in sentence context. *Journal of the Acoustical Society of America, 85*, 2135–2153.
Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America, 71*(6), 1596–1605.
Turk, A., & Sawusch, J. R. (1997). The domain of accentual lengthening in American English. *Journal of Phonetics, 25*, 25–41.
Turk, A., & Shattuck-Hufnagel, S. (2000). Word-boundary-related durational patterns in English. *Journal of Phonetics, 28*, 397–440.
Turk, A., & White, L. (1999). Structural effects on accentual lengthening in English. *Journal of Phonetics, 27*, 171–206.
Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 710–720.
Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38*, 133–149.
Welby, P. (2003). *The slaying of Lady Mondegreen, being a study of French tonal association and alignment and their role in speech segmentation. Ph.D. dissertation, The Ohio State University*.
Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50*, 351–360.
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.
Woodrow, H. (1909). A quantitative study of rhythm. *Archives of Psychology, 14*, 1–66.
Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review, 18*, 54–77.