Title
Distal prosody influences lexical interpretation in online sentence processing

Journal
Proceedings of the Annual Meeting of the Cognitive Science Society, 33(33)

ISSN
1069-7977

Authors
Brown, Meredith
Salverda, Anne Pier
Dilley, Laura C.
et al.

Publication Date
2011

Peer reviewed

Distal prosody influences lexical interpretation in online sentence processing

Meredith Brown (mbrown@bcs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Meliora Hall, Box 270268, Rochester, NY 14627-0268

Anne Pier Salverda (asalverda@bcs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Meliora Hall, Box 270268, Rochester, NY 14627-0268

Laura C. Dilley (ldilley@msu.edu)
Department of Communicative Sciences & Disorders, Michigan State University
116 Oyer, East Lansing, MI 48824

Michael K. Tanenhaus (mtan@bcs.rochester.edu)
Department of Brain & Cognitive Sciences, University of Rochester
Meliora Hall, Box 270268, Rochester, NY 14627-0268

Abstract

Previous work examining prosodic cues in online spoken word recognition has focused primarily on local cues to word identity. However, recent studies have suggested that sentence-level prosodic patterns can also influence the downstream interpretation of lexically ambiguous syllables (Dilley & McAuley, 2008; Dilley, Mattys, & Vinke, 2010). To test the hypothesis that these distal prosody effects are based on expectations about the organization of upcoming material, we conducted a visual world experiment using fixations to competing alternatives such as *pan* and *panda*, which differ in the presence or absence of a prosodic boundary after *pan(-)*. As predicted, the acoustic properties of distal sentence material affected the proportion of fixations to the monosyllabic competitor beginning 200 ms after the onset of the target word. These findings support the hypothesis that expectations based on perceived prosodic patterns in the distal context influence lexical segmentation and recognition.

Keywords: Prosody; spoken word recognition; lexical competition; perceptual organization; visual world paradigm

Introduction

Expectation and prediction are increasingly playing a major role in models of language processing (e.g., Levy, 2008; Jurafsky, 1996; Hale, 2001), and effects of expectations based on lexical and structural properties of sentences are well documented (reviewed in Kamide, 2008). However, another possible source of expectations that may be useful in language processing are global prosodic patterns, including prosodic phenomena that are distal to (i.e., several syllables removed from) the locus of processing. Such patterns may also influence the interpretation of proximal sentence material by contributing to listeners' expectations about the lexical structure of upcoming material. Most work at the interface of prosody and spoken word recognition to date has focused on the effects of local prosodic cues to word identity, such as stress, duration, and the strength of local prosodic boundaries (Cutler & van Donselaar, 2001; Salverda, Dahan, & McQueen, 2003; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004). In this paper, we provide evidence that distal prosody generates expectations based on perceived global prosodic patterns, and that these expectations constrain word segmentation and recognition.

Background

Previous work using non-linguistic auditory stimuli ranging from simple tone sequences to musical passages has shown that listeners detect pitch, temporal, and/or amplitude patterning in distal (i.e., non-local) auditory context. Such patterning influences the metrical and grouping structures that listeners perceive downstream (reviewed in Handel, 1989; Huron, 2006). For example, listeners tend to perceptually organize a continuous stream of tones of equal duration, temporal separation and amplitude that differ only in terms of pitch height (as high or low pitched) into repeating sequences of either high-low or low-high perceived groups, with the first element in each group perceived as accented (Woodrow, 1911; Thomassen, 1982). Although high tones are more frequently perceived as more prominent than low tones, the perceived relative prominence of high and low tones within each bitonal group depends on the context. To describe such perceptual proclivities in the domain of music processing, Lerdahl and Jackendoff (1983) proposed parallelism preference rules, which state that when a musical passage contains segments that are perceived as similar or repetitive, parallel parts of segments will be construed as having similar metrical and grouping structure. These studies suggest that distal regularities in the acoustic characteristics of auditory stimuli can influence the processing of proximal material.

Reports from linguistics suggest that prosodic context often exhibits regularities in pitch, duration, and/or amplitude. In particular, speech intonation and rhythm often show characteristics that listeners perceive as patterning (Couper-

Kuhlen, 1993; Dainora, 2001; Pierrehumbert, 2000). For example, listeners tend to hear stressed syllables as perceptually isochronous, i.e., as occurring at regular intervals (e.g., Lehiste, 1977), and speakers tend to use similar intonation patterns on accented syllables within an intonational phrase (Couper-Kuhlen, 1993; Crystal, 1969; Dainora, 2001; Pierrehumbert, 2000).

Based on these observations, Dilley and McAuley (2008) proposed that regularities in distal prosodic patterning might generate structural expectations in the domain of speech processing similar to the effects that have been observed in non-speech processing. According to the *perceptual grouping hypothesis*, distal prosodic patterns influence language processing by inducing periodic expectations about the organization of upcoming material, including the locations of major prosodic boundaries relative to upcoming syllables within a continuous speech stream. These prosodic expectations are in turn predicted to constrain the lexical candidates considered by listeners to words that are consistent with the anticipated prosodic structure.

In support of the perceptual grouping hypothesis, Dilley and McAuley (2008) showed that the segmentation of a sequence of lexically ambiguous syllables is sensitive to prosodic patterns across words distal to the ambiguous syllables. For example, in the sequence *channel dizzy foot-note-book-worm*, the final four syllables have lexically ambiguous structure. Manipulating the prosody of the first five syllables, i.e., *channel dizzy foot(-)*, influenced listeners' perception of the final three syllables in the sequence, i.e., *note-book-worm*. When asked to report the final word they heard in each sequence, participants reported more two-syllable compound words (i.e., *bookworm*) when the prosody across the distal context encouraged a grouping of the ambiguous syllables into two-syllable units. Likewise, participants given an incidental memory task after exposure to all sequences recognized a relatively high proportion of two-syllable words when the distal context predicted a two-syllable segmentation.

These findings have since been replicated and extended by Dilley, Mattys, and Vinke (2010), who used stimuli consisting of strings of lexical items ending in potentially end-embedded words (e.g., *turnip* vs. *nip*). They showed that distal prosody is a quite robust cue to lexical organization, even when semantic context or proximal acoustic cues suggested a different structure. Moreover, using a cross-modal identity priming paradigm, Dilley et al. (2010) showed that distal prosody affected the reaction time and accuracy of lexical decisions. These findings provide initial evidence of distal prosodic effects on early lexical processing consistent with the predictions of the perceptual grouping hypothesis, but several questions remain. One question concerns the time course of the observed distal prosody effects: If it is the case that the effects of distal prosodic patterns are indeed mediated by listeners' expectations, these effects would be expected to emerge in the earliest moments of processing. In addition, distal prosody effects have so far been demonstrated only in the processing of lists of words lacking grammatical structure. Thus, it is important to evaluate whether listeners are sensitive to distal prosodic patterns in more ecologically valid contexts, such as in the processing of sentences.

Experiment overview

The present study uses the visual world paradigm (Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995) to assess the effect of distal prosody on listeners' online comprehension of words such as *panda*, which are temporarily ambiguous due to the presence of an onset-embedded word (i.e., *pan*). In this experiment, the acoustic characteristics of sentence material distal to the target word are manipulated, while the acoustic characteristics of material proximal to the target word are held constant. If distal prosody influences listeners' expectations concerning the prosodic organization of upcoming lexical material, the prosodic context should affect listeners' perception of prosodic boundaries in the earliest moments of processing the target word. It should therefore be possible to observe greater or lesser lexical activation of the embedded competitor word (as indexed by fixations to the competitor image), depending on whether the distal prosodic context biases the listener to expect a major phrasal boundary after the embedded word.

Methods

Participants

Forty-three participants from the University of Rochester and the surrounding community consented to participate in the visual world experiment. All participants were native English speakers with normal hearing and corrected visual acuity.

Materials

The 20 speech stimuli used in the experiment were grammatical declarative sentences containing a target word with an onset-embedded competitor word (e.g., *panda*). Target words were chosen such that both the word itself and the embedded competitor word were highly imageable. The distal context preceding the target word in each stimulus consisted of two disyllabic words with initial primary stress, followed by one monosyllabic word (e.g., *Heidi sometimes saw*). Likewise, the proximal context preceding the target word consisted of an additional monosyllabic word (e.g., *that*) followed by the target word itself (e.g., *panda*). The sentence material following the target word consisted of a 3-5 syllable continuation. An example stimulus is shown in Figure 1.

To reduce the probability of participants noticing a tendency for target words to be the longer of two phonologically related words, 20 filler items were included in which the target word was the shorter of two phonologically related words corresponding to items in the visual display. In addition, 20 filler items were included in which two pictures with phonologically related labels appeared in the display, but the target word did not overlap phonologically with either word.

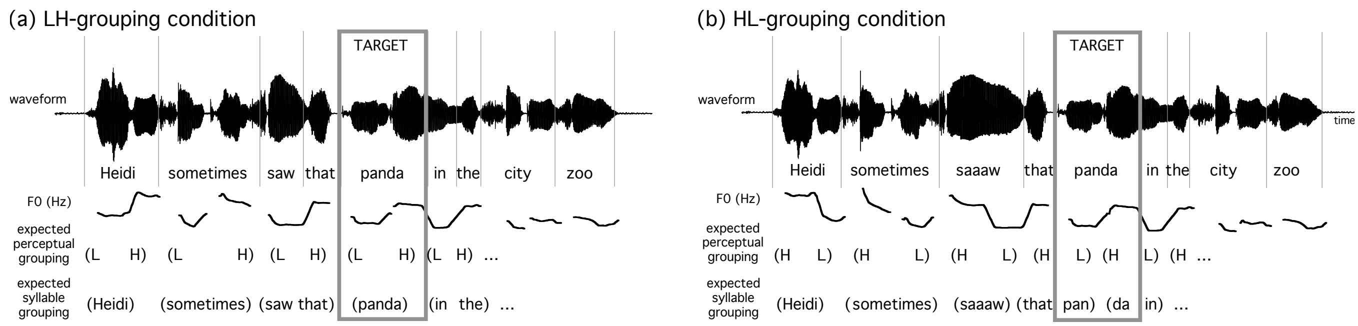

Figure 1: Explanation of the resynthesis methods performed to generate stimuli in (a) the LH-grouping condition, and (b) the HL-grouping condition. The fundamental frequency (F0) of the first five syllables and the duration of the fifth syllable were manipulated to encourage or discourage the perception of a phonological phrase boundary following the embedded word *pan*. The acoustic properties of the stimuli were held constant from the onset of the sixth syllable through the end of the sentence.

The structure of the distal context within the filler items (e.g., the number of syllables) was varied to reduce the information available to participants about the likely sentence position of the target word in upcoming experimental items.

The first author recorded all items as WAV files using a Marantz PMD 660 digital recorder in a sound-attenuated booth. The sentences were produced with minimal F0 excursions and slight F0 declination across each utterance. The pitch synchronous overlap-and-add (PSOLA) algorithm (Moulines & Charpentier, 1990) was then used in Praat (Boersma & Weenink, 2010) to create two resynthesized versions of each recording. Resynthesizing the stimuli allowed us to create auditory materials that were relatively natural-sounding, yet carefully controlled.

The distal F0 manipulation involved shifting all pitch points within the vocalic portion of each context syllable up by 35 Hz (for high syllables) or down by 25 Hz (for low syllables). This manipulation allowed us to maintain the natural pitch declination of the original recording while imposing periodic alternations onto the F0 contour. The F0 contours across the non-vocalic portions of the speech were determined by monotonic interpolation between the F0 value at the end of the previous vocalic region and the F0 value at the start of the subsequent vocalic region.

In the low-high (LH) grouping condition, phonological phrase boundaries in the distal context were temporally aligned with the edges of roughly isochronous sequences of low and high syllables (Figure 1a). This timing was achieved by modifying the duration of the fifth rime such that the fifth intervocalic interval (i.e., the interval between vowel onsets in the fifth and sixth syllables) was equivalent to the mean durations of the sixth and seventh intervocalic intervals (cf. Dilley & McAuley, 2008; Dilley et al., 2010). The LH-grouping distal context was predicted to facilitate the perception of prosodic boundaries preceding and following the target word, consistent with either a *pan* or a *panda* segmentation.

In the high-low (HL) grouping condition, on the other hand, prosodic boundaries in the distal context were aligned with the edges of sequences of high and low tones (Figure 1b). In this condition, the duration of the fifth syllable was lengthened in Praat such that the total duration of the fifth intervocalic interval was equivalent to the mean value of the sum of the first two intervocalic intervals and the sum of the third and fourth intervocalic intervals. This duration manipulation enabled the fifth syllable to be paired with two F0 levels instead of one while maintaining the approximately isochronous timing of each high-low F0 sequence. The HL-grouping distal context was predicted to bias listeners to perceive the target word as two prosodic units – e.g., *pan*)(*da* – with a boundary following *pan*. This perceived prosodic boundary was expected to increase the relative activation of the embedded word, and consequently the proportion of fixations to the competitor image.

Importantly, the acoustic characteristics of the sentence were held constant across both conditions from the start of the syllable preceding the target word (i.e., the start of the proximal context) through the end of the sentence (Figure 1). In both the LH- and HL-grouping conditions, the word preceding the target word had a high tone and the target word itself had a LH tonal sequence. Controlling the acoustic characteristics of the target word and proximal context ensured that any differences in participants' fixation patterns between the LH- and HL-grouping conditions could not be attributed to proximal prosodic cues.

Pitch and duration manipulations were performed on the filler items as well. To eliminate a consistent association between pitch and duration patterns, half of the filler items beginning with HL F0 contours contained one shortened syllable at a variable position in the sentence, whereas the other half contained one lengthened syllable. Likewise, half of the filler items starting with LH F0 contours contained one shortened syllable, and the other half had one lengthened syllable.

Procedure

Each trial began with the presentation of a visual display containing four colored clip-art images, two of which corresponded to the target word and the competitor word (Figure 2). The remaining two distractor images were selected such that they were visually distinct from the target and competitor images and their labels were phonologically unrelated to the target word. After 500 ms of visual preview, the participant heard a stimulus over Sennheiser HD 570 headphones and was asked to click on the picture that was referred to in the spoken sentence. They were not given feedback on their performance during the experiment. Pilot testing indicated that participants generally noticed that the speech had been manipulated, so we employed the cover story that the goal of the present study was to evaluate the comprehensibility of synthesized speech stimuli for use in another study. Throughout the study, eye movements were tracked and recorded using a head-mounted SR Research EyeLink II system sampling at 250 Hz. Drift correction procedures were performed after every fifth trial.

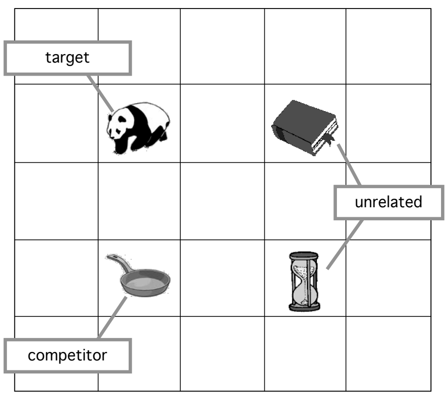

Figure 2: An example display, with picture labels added.

Four lists were constructed by randomizing the positions of the the images on the screen within each trial and pseudo-randomizing the order of trials within the list. Within each list, half of the experimental trials had HL-grouping prosody and half had LH-grouping prosody. The assignment of items to each type of grouping was counterbalanced across participants, for a total of eight lists. Four practice trials were included to familiarize participants with the study procedure.

Results

Three participants were excluded from analysis because they did not meet inclusion criteria or did not complete the whole experiment. These participants were replaced such that analyses of fixation patterns included data from five participants for each of the eight experiment lists. Overall, participants found the referent identification task to be easy, and clicked on the incorrect picture on less than 3% of all experimental trials. These trials were excluded from further analysis.

Figure 3 shows the mean proportion of fixations to the target, competitor, and distractor images as a function of condition starting at the onset of the target word. In the LH-grouping condition, the proportions of fixations to the target and competitor images were roughly equivalent until approximately 200 ms following the offset of the embedded word (566 ms), at which time participants began to converge on the target image. However, in the HL-grouping condition, fixation patterns revealed a competitor bias that emerged at around 200 ms following target onset (i.e., approximately the time at which signal-driven fixations would be expected to emerge) and persisted until approximately 200 ms following the offset of the embedded word.

For statistical analysis of these fixation curves, we employed growth curve analysis (GCA; Mirman et al., 2008), a technique that is increasingly being applied to visual world data. GCA is a variant of multi-level regression modeling in which orthogonal power polynomial terms are fit to time-series data to model variations in curve parameters (e.g., the intercept and slope) that can be attributed to the independent variable(s) and/or to individual differences in processing. Unlike other statistical methods traditionally used to evaluate effects of experimental factors in the visual world paradigm (e.g., analysis of variance for proportions of fixations averaged over a single time window, or repeated-measures analyses across successive time windows), GCA explicitly models the trajectory of change over time through the comparison of curve parameters. This approach avoids several disadvantages of standard analysis of variance approaches, including the violation of independence assumptions and the loss of fine-grained temporal detail. GCA is also an appropriate analysis technique for this experiment because previous psycholinguistic research has revealed individual differences in the effects of prosody on the interpretation of spoken language (e.g., Fraundorf & Watson, 2010). GCA, unlike standard analyses of variance, takes individual variation into account by explicitly incorporating parameters that estimate the effects of factors for each participant.

Growth curve analyses were conducted separately for target and competitor fixations and included data from 150-1200 ms after the onset of the target word (i.e., from the earliest point at which signal-driven fixations could theoretically be observed until the approximate time at which the proportion of target fixations reached its maximum value). Separate analyses were done to assess by-subject and by-item effects (indicated below as B_1 and B_2, respectively). The perceptual grouping hypothesis most directly predicts differences in the intercept term, which represents the average height of each curve, though any differences in other curve parameters between conditions are attributable solely to the distal context and are therefore of interest.

For competitor fixations, cubic and quartic terms were included in addition to intercept and slope terms, due to the characteristic presence of three inflection points in the competitor fixation curves. Distal prosody condition had a significant effect on the intercept term (B_1 = -.031, t_1 = -2.911, $p < .01$; B_2 = -.031, t_2 = -2.810, $p < .05$), indicating that the mean proportion of fixations to the competitor was significantly higher in the HL-grouping condition than in the LH-

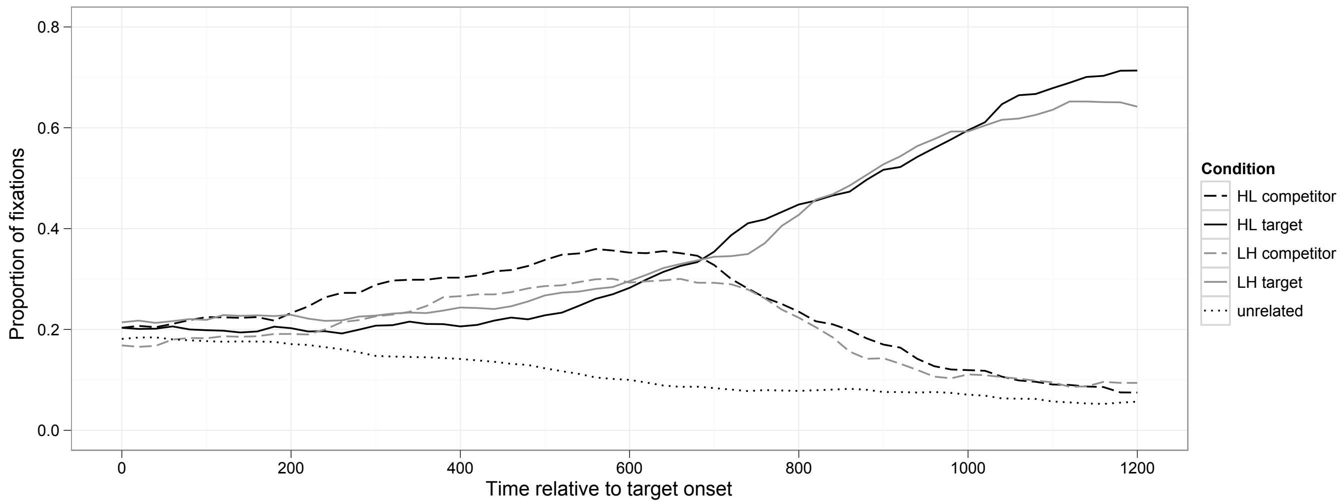

Figure 3: Proportion of fixations to the target, competitor, and unrelated images in the HL-grouping and LH-grouping conditions. The mean duration of the embedded word was 366 ms, and the mean duration of the target word was 675 ms.

grouping condition. The effect of condition on the slope term, on the other hand, was not significant (B_1 = .064, t_1 = .673, $p > .1$; B_2 = .079, t_2 = .730, $p > .1$). This result reflected a similar rate of change in fixation proportions in the HL-grouping and LH-grouping conditions. A significant effect of condition on the quadratic term (B_1 = -.224, t_1 = -11.331, $p < .0001$; B_2 = -.226, t_2 = -12.006, $p < .0001$) revealed that the steepness of the curve preceding and following the point of peak competitor fixations was greater in the HL-grouping condition than in the LH-grouping condition. Likewise, condition had a significant effect on the cubic term (B_1 = -.205, t_1 = -10.379, $p < .0001$; B_2 = -.207, t_2 = -11.025, $p < .0001$) but not the quartic term (B_1 = -.022, t_1 = -1.099, $p > .1$; B_2 = -.013, t_2 = -.695, $p > .1$). These terms also pertain to the shape of the curves around the inflection points, but are otherwise difficult to interpret (Mirman et al., 2008). In summary, the competitor fixation curve had a higher mean value and a steeper curve surrounding inflection points in the HL-grouping condition than in the LH-grouping condition.

For target fixations, distal prosody condition did not have a significant effect on either the intercept term (B_1 = .003, t_1 = .257, $p > .1$; B_2 = .001, t_2 = .085, $p > .1$) or the slope term (B_1 = -.106, t_1 = -.931, $p > .1$; B_2 = -.118, t_2 = -1.095, $p > .1$). However, condition did have a significant effect on the quadratic term (B_1 = .235, t_1 = 8.124, $p < .0001$; B_2 = .247, t_2 = 7.320, $p < .0001$), and on the cubic term (B_1 = .080, t_1 = 2.746, $p < .01$; B_2 = .085, t_2 = 2.508, $p < .05$). These results suggest that the steepness of the curve around the inflection points was greater in the LH-grouping condition than in the HL-grouping condition, but otherwise the target fixation curves did not differ significantly across conditions.

Discussion

According to the perceptual grouping hypothesis, distal prosodic patterns can influence online language processing by giving rise to periodic expectations about the prosodic organization of upcoming material. The perceptual grouping hypothesis predicts greater lexical activation of words whose boundaries are consistent with the expected prosodic structure, in the earliest moments of processing such words. We tested this prediction in a visual world study by comparing fixations to competing lexical alternatives such as *pan* and *panda*. Results supported the predictions of the perceptual grouping hypothesis. Starting around 200 ms after the onset of the target word, the proportion of competitor fixations was higher overall when the distal prosody was consistent with a phonological phrase boundary following *pan-* than when the distal prosody was consistent with prosodic boundaries preceding and following *panda*. This pattern of results suggests that the embedded competitor word was more highly activated in the former condition immediately following the onset of *panda*. The observed differences in the proportion of competitor fixations across conditions cannot be attributed to the acoustic characteristics of the target word or its proximal context, since the acoustic-phonetic properties of the proximal context were identical across conditions. We therefore conclude that distal prosodic patterns can influence the prosodic structure that listeners project forward onto upcoming material in the speech stream, and that these prosodic expectations can interact with online lexical segmentation and access.

These findings are congruent with studies showing that other types of distal prosody influence word recognition. For example, the distal speaking rate has been shown to influence the segmentation of lexically ambiguous syllables in Dutch (Reinisch, Jesse, & McQueen, in press) and the perception of function words in English (Dilley & Pitt, 2010). Speech processing can also be influenced by non-linguistic pitch information in the preceding context: Distal sequences of sine wave tones with different spectral properties can influence speech categorization (Holt, 2005). Taken together with these findings, the results from our study support the notion that the global prosodic context influences not only the interpretation of local prosodic cues, but also lexical segmentation and recognition processes.

An important remaining question concerns the characteristics of the distal prosodic patterns that are relevant for online speech comprehension. In our stimuli, the distal prosody contained multiple regularities that could potentially contribute to listeners' prosodic expectations. For example, listeners may be sensitive to the regular co-occurrence of the edges of prosodic constituents and the edges of LH or HL tone groups in the distal context. Alternatively, because the distal context consistently employed alternating sequences of strong and weak syllables, listeners may primarily be sensitive to the co-occurrence of certain types of tones with metrically prominent syllables. Work in progress is assessing the effects of distal prosody on the comprehension of words with different stress patterns to adjudicate between these possibilities.

Acknowledgments

This research was supported by a Javits fellowship and NSF predoctoral fellowship to MB, NSF grant BCS-0847653 to LCD, and NIH grants HD27206 and DC0005071 to MKT. We gratefully acknowledge Dana Subik for assistance with participant recruitment and testing, and the Tanenhaus lab and the audience at AMLaP 2010 for helpful discussions.

References

Boersma, P., & Weenink, D. (2010). Praat: doing phonetics by computer [Computer program]. Version 5.1.25, retrieved 20 January 2010 from http://www.praat.org.

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.

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

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

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

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

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

Dilley, L., & McAuley, J. D. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language*, *59*, 291–311.

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

Fraundorf, S. H., & Watson, D. G. (2010). Who cares about prosody? Predicting individual differences in sensitivity to pitch accent in online reference resolution. Poster presented at Architectures and Mechanisms for Language Processing 2010, York, UK.

Hale, J. (2001). A probabilistic Earley parser as a psycholinguistic model. *Proceedings of the 2nd Conference of the North American Chapter of the Association for Computational Linguistics*, *2*, 159–166.

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

Holt, L. L. (2005). Temporally nonadjacent nonlinguistic sounds affect speech categorization. *Psychological Science*, *16*, 305–312.

Huron, D. (2006). *Sweet anticipation: Music and the psychology of expectation.* Cambridge, MA: MIT Press.

Jurafsky, D. (1996). A probabilistic model of lexical and syntactic access and disambiguation. *Cognitive Science*, *20*, 137–194.

Kamide, Y. (2008). Anticipatory processes in sentence processing. *Language and Linguistics Compass*, *2*, 647–670.

Lehiste, I. (1977). Isochrony reconsidered. *Journal of Phonetics*, *5*, 253–263.

Lerdahl, F., & Jackendoff, R. (1983). *A generative theory of tonal music.* Cambridge, MA: MIT Press.

Levy, R. (2008). Expectation-based syntactic comprehension. *Cognition*, *106*, 1126–1177.

Mirman, D., Dixon, J. A., & Magnuson, J. S. (2008). Statistical and computational models of the visual world paradigm: Growth curves and individual differences. *Journal of Memory and Language*, *59*, 475–494.

Moulines, E., & Charpentier, F. (2008). Pitch-synchronous waveform processing techniques for text-to-speech synthesis using diphones. *Speech Communication*, *9*, 666–688.

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

Reinisch, E., Jesse, A., & McQueen, J. M. (in press). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance*, *105*, 466–476.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, *90*, 51–89.

Tanenhaus, M. K., Spivey-Knowlton, M., Eberhard, K., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632–1634.

Thomassen, J. M. (1982). Melodic accent: Experiments and a tentative model. *Journal of the Acoustical Society of America*, *71*, 1596–1605.

Woodrow, H. (1911). The role of pitch in rhythm. *Psychological Review*, *18*, 54–77.