Journal of Phonetics 46 (2014) 68–85

Contents lists available at ScienceDirect

Journal of Phonetics

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

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

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

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

ARTICLE INFO

Article history:
Received 16 September 2013
Received in revised form
12 March 2014
Accepted 16 June 2014

Keywords:
Distal prosody
Word segmentation
Perceptual grouping
Intonation

ABSTRACT

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

1. Introduction

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

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

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

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

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

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

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

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

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