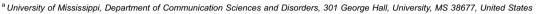Research Article

Finding word boundaries in Indian English-accented speech

Kara Hawthorne [a,b,*], Juhani Järvikivi [b], Benjamin V. Tucker [b]

[a] University of Mississippi, Department of Communication Sciences and Disorders, 301 George Hall, University, MS 38677, United States
[b] University of Alberta, Department of Linguistics, 4-32 Assiniboia Hall, Edmonton, AB T6G 2E7, Canada

ARTICLE INFO

ABSTRACT

The majority of English nouns, verbs, and adjectives begin with a stressed syllable, and listeners exploit this tendency to help parse the continuous stream of speech into individual words. However, the acoustic manifestation of stress depends on the variety of English being spoken. In two visual world eye-tracking experiments, we tested if Indian English-accented speech causes Canadian English listeners to make stress-based segmentation errors. Participants heard Canadian- or Indian-accented trisyllabic sequences that could be segmented in two ways, depending on the perceived location of stress. For example, [hæ.pi.tsə] could be segmented as happy/[tsə] if it is perceived to have stress on the first syllable or as [hæ]/pizza if it is perceived to have stress on the second syllable. Results suggest that Indian English-accented speech impairs segmentation in Canadian listeners, and that both accented pitch and other features of the Indian English accent contribute to segmentation difficulties. Findings are interpreted with respect to models of how similarity between two languages impacts the listener's ability to segment words from the speech stream.

© 2017 Elsevier Ltd. All rights reserved.

1. Introduction

In written language, word boundaries are clearly indicated by the presence of a space, but no space or pause delineates most words in spoken language. Instead, listeners must rely on other cues to segment a continuous acoustic signal into discrete words. Adults listening to their native language have the advantage of extensive lexical and grammatical knowledge, which they can use to map portions of the speech stream onto known lexical items with the aid of top-down expectations (McClelland & Elman, 1986). In addition, prosody, phonotactics, co-articulation, and allophonic variants serve as useful bottom-up cues (Davis, Marslen-Wilson, & Gaskell, 2002; Mattys, 2004; Mattys, White, & Melhorn, 2005; Ordin & Nespor, 2013). However, bottom-up acoustic cues are subject to cross-speaker variation, including variability due to the speaker's regional or foreign accent. In this paper, we focus on the impact of accented prosody[1] – stress, in particular – on segmenting speech into words. More specifically, we ask if

Indian English, in which stressed syllables are manifested similarly but not identically to many other varieties of English, impairs word segmentation for native speakers of Canadian English. We discuss our results in light of related work on the relationship between second language learning outcomes and similarity of the native and second languages: we extend these models to examine the impacts of similarity across accents.

1.1. Prosody and word segmentation

There are two ways that prosody helps listeners segment speech into words. First, across languages, boundaries of prosodic phrases always coincide with boundaries of lower-level prosodic constituents. For example, the end of an utterance is also always the end of a phonological phrase, as well as the end of a word. Prosodic phrase boundaries are typically marked with lengthening of final syllables, and listeners across ages and languages use final lengthening as a signal that they have reached the end of a word (Christophe, Gout, Peperkamps, & Morgan, 2003; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Endress & Hauser, 2010; Shukla, Nespor, & Mehler, 2007).

Second, lexical-level prosody can serve as a reliable segmentation cue, though it varies significantly from language to language. In English, 85% of content word tokens begin with

* Corresponding author at: University of Mississippi, Department of Communication Sciences and Disorders, 301 George Hall, University, MS 38677, United States.
E-mail addresses: khawthor@olemiss.edu (K. Hawthorne), jarvikiv@ualberta.ca (J. Järvikivi), bvtucker@ualberta.ca (B.V. Tucker).
[1] In this paper, we will use the term "accent" to refer to a regional or foreign accent. When we discuss pitch accents, we will specify them as such.

a stressed syllable (Cutler & Carter, 1987), so a reasonable first-pass segmentation strategy is to place a word boundary before each syllable that is perceived as stressed. There is an extensive literature supporting the role of stress in word segmentation, as well as lexical access and infant word learning, in English and other languages with lexical stress (e.g., Cutler & Norris, 1988; Grosjean & Gee, 1987; Johnson & Jusczyk, 2001; Jusczyk, Cutler, & Redanz, 1993). Stress is particularly helpful for segmentation when other sources of information – sentential context, lexical knowledge, and segmental cues – are unreliable (Mattys et al., 2005). For example, stress is useful for segmenting speech in noise (Mattys, 2004; Smith, Cutler, Butterfield, & Nimmo-Smith, 1989) or produced by a speaker with dysarthria (Liss, Spitzer, Caviness, Adler, & Edwards, 1998). Other types of lexical-level prosody help segment words in languages that do not have lexical stress, such as French (e.g., Michelas & D'Imperio, 2010; Tremblay, Coughlin, Bahler, & Gaillard, 2012), Korean (e.g., Kim & Cho, 2009) and Japanese (Otake, Hatano, Cutler, & Mehler, 1993).

Prosody is an important segmentation cue for adults acquiring a second language (L2). Final syllable lengthening – potentially a universal cue to prosodic boundaries – is reliably used for L2 segmentation by learners with a variety of language backgrounds (e.g., Endress & Hauser, 2010; Kim, Broersma, & Cho, 2012; Ordin & Nespor, 2013, 2016). On the other hand, lexical stress is a less reliable L2 segmentation cue. Not all languages have lexical stress, and those that do vary in how stress is manifested acoustically and how it correlates with word boundaries. In fact, previous research has found that listeners use stress and other lexical-level prosodic segmentation strategies from their native language (L1) when listening to a natural or artificial L2. They do so even when those strategies are not appropriate for the target language (Cutler, Mehler, Norris, & Segui, 1986, 1992; Otake et al., 1993; Vroomen, Tuomainen, & de Gelder, 1998).

Recent work has attempted to determine the situations in which lexical-level prosody can facilitate or hinder segmentation for an L2 learner, primarily by manipulating the location of prosodic prominence in artificial grammar experiments. Tyler and Cutler (2009) found that English speakers, who are used to word-initial prominence, can segment more successfully when word-initial syllables are marked by pitch, whereas French speakers, who are used to word-final prominence, are more successful at segmenting when word-final syllables are marked with pitch. Likewise, Kim et al. (2012) report that Korean speakers use F0 (fundamental frequency) rise on word-final syllables to segment an artificial grammar, while Dutch speakers do not. Again, this would be expected, based on the prosodies of Dutch and Korean. Ordin and Nespor (2013, 2016) similarly found that segmentation was disrupted for Italian and German speakers when pitch or lengthening cues conflicted with the typical location of stressed syllables in the learner's L1. Together, these studies suggest that L2 segmentation is impaired when prosodic prominence occurs in an unexpected location.

Tremblay, Broersma, Coughlin, and Choi (2016) argue that the difficulty of learning to segment an L2 depends not only on *where* prominence is marked in the two languages, but also on *how*. They report that both Korean and English L2 learners of French segment French differently than native speakers, but that the Korean speakers had a harder time using F0 as a cue to word-final boundaries than English speakers. Like French, Korean uses rising F0 as a cue to word-final, accentual phrase-final boundaries, though the precise alignment of the pitch cue is different in the two languages. English, on the other hand, marks word-initial boundaries using stress, which is cued in part by rising F0 (see Section 1.2).

Tremblay and colleagues propose the *Prosodic-Learning Interference Hypothesis (PLIH)* to account for these results. This hypothesis states that acquisition of a new segmentation cue is more difficult if the L1 and L2 prosodic systems are similar (e.g., French and Korean), requiring learning of subtle distinctions, rather than dramatically different (e.g., French and English in their experiment, as well as most of the language pairings in related literature). The authors define prosodic similarity as when a similar prosodic cue marks the same word boundary (word-initial or word-final), but not in an identical way (or there would be nothing to learn). For example, French and Korean are prosodically similar because they both mark word-final boundaries with rising F0, but the pitch cue is aligned differently in the two languages. The PLIH builds on models of nonnative perception of consonants and phonemes. For example, the *Speech Learning Model* (Flege, 1995; see also the Perceptual Assimilation Model: Best & Tyler, 2007) proposes that novel L2 phonemes are more difficult to learn when the learner's L1 has a similar phoneme than if the new sound is completely novel.

Prosody varies, not only across languages, but also across foreign- and regionally-accented varieties of the same language. These prosodic differences impact the perception of accentedness and intelligibility (Boula de Mareuil & Vieru-Dimulescu, 2006; Jilka, 2000; Polyanskaya, Ordin, & Busa, 2016; Quené & Van Delft, 2010; Ulbrich & Mennen, 2016; Winter & O'Brien, 2013). While lexical-level prosodic variation triggers segmentation challenges for listeners in their L2, it is unknown whether similar challenges occur for listeners in their L1 when the talker is non-native or has an unfamiliar regional accent.

Likewise, while L1-L2 prosodic similarity can cause difficulties in L2 acquisition, it is unclear if similarity helps or hinders word segmentation when two people speak different varieties of the same language. Adapting to an unfamiliar accent poses fundamentally different challenges from learning an L2. An L2 learner is consciously attempting to acquire a new linguistic system, while an individual listening to a person with an unfamiliar accent is learning to map a set of unexpected speech features to a natively-acquired language. If the PLIH holds for accent adaptation, any initial segmentation difficulties caused by prosodic dissimilarities between the listener's and talker's accents should be mitigated by learning (typically referred to as "adaptation" in the accent literature) if the dissimilarities are significant or if the two varieties mark different word boundaries. On the other hand, difficulties should persist if the two varieties are prosodically similar, with relatively small differences that require listeners to adapt to subtle changes.

No previous work has looked at how foreign or regionally accented stress impacts word segmentation, but there is some evidence that accented prosody impairs the higher-level process of lexical access. Reinisch and Weber (2012) tested