Journal of Memory and Language 66 (2012) 509–529

Contents lists available at SciVerse ScienceDirect

Journal of Memory and Language

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

How does context play *a part* in splitting words *apart*? Production and perception of word boundaries in casual speech

Dahee Kim [a,*], Joseph D.W. Stephens [b], Mark A. Pitt [c]

[a] Department of Linguistics, Ohio State University, Columbus, OH, United States
[b] Department of Psychology, North Carolina A&T State University, Greensboro, NC, United States
[c] Department of Psychology, Ohio State University, Columbus, OH, United States

ARTICLE INFO

Article history:
Received 30 April 2010
revision received 10 December 2011
Available online 26 January 2012

Keywords:
Word segmentation
Casual speech
Speech production
Speech perception

ABSTRACT

Four experiments examined listeners' segmentation of ambiguous schwa-initial sequences (e.g., *a long* vs. *along*) in casual speech, where acoustic cues can be unclear, possibly increasing reliance on contextual information to resolve the ambiguity. In Experiment 1, acoustic analyses of talkers' productions showed that the one-word and two-word versions were produced almost identically, regardless of the preceding sentential context (biased or neutral). These tokens were then used in three listening experiments, whose results confirmed the lack of local acoustic cues for disambiguating the interpretation, and the dominance of sentential context in parsing. Findings speak to the H&H theory of speech production (Lindblom, 1990), demonstrate that context alone guides parsing when acoustic cues to word boundaries are absent, and demonstrate how knowledge of how talkers speak can contribute to an understanding of how words are segmented.

Introduction

Spoken language is often a continuous stream of speech. For comprehension to succeed, the listener must segment this stream into a sequence of individual words. A substantial literature has been devoted to determining the degree to which information about word-boundary locations is present in the acoustics of speech (Lehiste, 1960), or dependent upon higher-order contextual factors such as the listener's interpretation of word meaning and sentence structure (Cole, Jakimik, & Cooper, 1980). By identifying the different sources of information, their relative importance, and how they are used in combination (Mattys, White, & Melhorn, 2005; Norris, McQueen, Cutler, & Butterfield, 1997), it is thought that a comprehensive theory of word segmentation can be constructed.

The purpose of the current study is to suggest that a consideration of the talker can inform our understanding of how spoken words are segmented. Successful communication requires the listener to have learned how to segment speech from a variety of talkers (e.g., native, foreign-accented) speaking in a variety of styles (e.g., careful vs. casual speech), and for talkers to have learned to speak with enough clarity so that listeners can parse the speech stream and comprehend the message. An understanding of how talkers actually speak is thus a potentially useful source of information for describing the segmentation problem and for addressing theoretical issues about segmentation in both production and perception. We pursue this idea by examining which acoustic cues to segmentation talkers provide in high-frequency, casually-produced utterances, and how this knowledge can influence thinking about solutions to the segmentation problem.

Previous studies have shown that spoken language is rich in acoustic cues to word boundaries. Major findings from the literature include lengthening of word-initial and word final segments and syllables (Beckman & Edwards, 1990; Lehiste, 1960) and shortening of segments and syllables that are not adjacent to a word boundary (Harris & Umeda, 1974; Klatt, 1976; Lehiste, 1972; Oller,

* Corresponding author. Fax: +1 614 292 8833.
E-mail addresses: Kim.2245@osu.edu (D. Kim), pitt.2@osu.edu (M.A. Pitt).
URL: http://ling.osu.edu/~daheekim (D. Kim).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2011.12.007

1973). For instance, Lehiste (1960) had speakers produce English words and word sequences with boundary ambiguities (e.g., *Grade A* vs. *gray day*; *nitrate* vs. *night-rate* vs. *Nye trait*) in isolation and embedded in short sentences. Duration and spectrographic analyses revealed a consistent and considerable durational difference among word-initial, word-medial, and word-final segments, in that word-initial or word-final segments were longer than word-medial segments. Related to the lengthening of word-initial segments, articulatory strengthening that occurs at the initial position of a word or a prosodic boundary has also been extensively studied (Byrd & Saltzman, 1998; Cooper, 1991). Byrd and Saltzman (1998), for instance, compared lip movements of syllable onset /m/ in different boundary conditions (e.g., word medial *mommamia* vs. word initial *Momma–Mimi* vs. across a word boundary *Momma, Mimi*, among others), and reported that lip movements become slower when /m/ is adjacent to a word or prosodic boundary than when it is not adjacent to any boundaries. Byrd, Kaun, Narayanan, and Saltzman (2000) extended this finding by confirming that articulatory gestures "get larger, longer, and further apart" when adjacent to a boundary. Additional phonetic properties that have been reported to correlate with word boundaries include amplitude contour, allophonic realizations of segments (e.g., clear /l/ vs. dark /l/; Umeda & Coker, 1975), spectral differences of vowels (Hoard, 1966; Lehiste, 1960), and degree of coarticulation (Krakow, 1989; Redford, Davis, & Miikkulainen, 2004). Taken together, the presence of such systematic variation in speech production demonstrates that talkers readily produce word boundary cues for listeners as they speak.

Although it is uncontroversial that acoustic cues to word boundaries can be found in the speech signal, relatively little is known about whether talkers produce these cues in utterances that are likely to occur in everyday speech, and whether talkers adjust their production of these cues based on the presence or absence of linguistic context making the message more ambiguous (e.g., *The baker looked at the drawing of a plump eye* vs. *plum pie*; Lieberman, 1963; Mattys & Melhorn, 2007).

In the Hypo- and Hyper-articulation (H&H) theory of speech production, Lindblom (1990) argues that talkers adapt their speaking style to the communicative demands of the listeners. More specifically, the theory states that talkers adjust their articulatory effort and its corresponding clarity of speech according to their estimate of how difficult it is for the listener to comprehend the message. If talkers estimate that comprehension would be difficult, due to the lack of strong contextual cues or background noise for instance, they would produce hyper-articulated speech to ensure that listeners comprehend the speech. If talkers believe the message is clear and unambiguous, they would default to produce hypo-articulated speech. The consequences for word segmentation are that talkers may produce clear acoustic cues to word boundaries when they estimate that listeners require clarity, but provide far weaker boundary cues in other communicative situations.

The preceding discussion makes it clear that consideration of the talker has implications for theorizing about how listeners segment words. Most notably, if the talker is responsible for producing speech at a minimally sufficient level of clarity, as suggested by H&H theory, the communication system could be placing an undue burden on the talker. A talker may cause communication to break down, for instance, by misestimating the level of clarity the listener requires. Such failures may be prevented if the perceptual system is robust against a high degree of uncertainty and ambiguity in the acoustic signal, by being less reliant on the talker speaking clearly. To ensure successful communication, a flexible and lenient model of speech perception may be preferred. However, such a model risks making the talker irrelevant. A theory of speech perception and word segmentation can be informed by knowledge about how talkers speak, including what acoustic cues are consistently present in the speech produced by the talker. For example, if aspiration in voiceless stops is shown to occur only at word onsets, there is a good reason for listeners to segment upon hearing it. Such reliable behavior of the talker can justify its specification in a model of the listener.

Studies of how listeners segment words demonstrate that they are efficient in exploiting all available information, acoustic and contextual. Acoustic cues that have been shown to affect listeners' segmentation include the acoustics of word-onset segments (Nakatani & Dukes, 1977), allophonic variation (Christie, 1974; Gow & Gordon, 1995), fundamental frequency contour (Ladd & Schepman, 2003), amplitude contour (Lehiste, 1960), and durational patterns corresponding to prosodic boundaries, including word boundaries (Cho, McQueen, & Cox, 2007; Salverda et al., 2007).

For example, Gow and Gordon (1995) showed that listeners are sensitive to the phonetic details distinguishing phonemically identical segments at different locations within a word. Using cross-modal semantic priming, they found that listeners were faster in deciding that *kiss* is a word after hearing the phrase *two lips* than after hearing the word *tulips*, suggesting that listeners are sensitive to the phonetic difference between the two primes. Studies focusing on the processing of embedded words have reported similar findings. Davis, Marslen-Wilson, and Gaskell (2002) found greater lexical activation for monosyllabic words such as *cap* when the syllable *cap* was originally produced as a whole word (e.g., in the phrase *cap tucked*) rather than as the initial syllable of a longer word (e.g., *captain*), which later was also confirmed by eye-tracking experiments (Salverda, Dahan, & McQueen, 2003).

Contextual cues that have been shown to affect listeners' segmentation include the lexical status of words in the stimuli (Mattys et al., 2005), the plausibility of possible interpretations of ambiguous word sequences (Mattys & Melhorn, 2007; e.g., *plump eye* or *plum pie* after hearing *The baker looked at the drawing of a* ... vs. *The surgeon looked at the drawing of a* ...), phonotactic constraints (McQueen, 1998), phonological properties of the language (Cutler & Norris, 1988; Norris et al., 1997), and the preceding sentential context (Cole et al., 1980). For example, Cole et al. (1980) showed that the semantic content of a preceding narrative influences segmentation. They had listeners detect a mispronunciation as they heard the sentence *they saw the carko on the ferry*. The sentence occurred