Lexical retrieval in read speech: The effect of speech prosody

Vahid Sadeghi[1], Negin Mohammadi Nafchi[2]

1. Introduction

For spoken information transfer to be successful, the listener needs to understand the meaning of the spoken utterance, thus, the message. The message a listener receives is encoded in an acoustic signal, which results from the physiological movements involved in speech production. The process of listening begins once this signal reaches the ear. After the initial psychoacoustic processing of the input, the listener separates speech from other sensory input that might reach the ear (see Bregman: 1990, for a review). The acoustic signal is sent to the auditory cortex via the auditory nerves and is then converted into an abstract representation used to access the mental lexicon, the stored representations of words. The next processing stage is called word recognition. At this stage, the listener has to segment the signal into meaningful discrete units. Once the words are recognized, the following processing stages are concerned with integration: listeners determine the syntactic and semantic properties of individual words and the syntactic and semantic relationships among them, and use this knowledge as well as pragmatic and world knowledge to understand and interpret the utterance.

As such, listeners employ a variety of linguistic patterns, including phonological, morphological, syntactic, and semantic ones to identify word boundaries in speech recognition. Based on the Metrical Segmentation Hypothesis, when linguistic information is not available, listeners use patterns of prosodic variation to identify word boundaries.

As proposed by Anne Cutler and others (Cutler & Butterfield, 1992; Cutler & Norris, 1988), the metrical segmentation strategy (hereafter MSS), relies on the fact that the majority of content words in stress-timed languages like English have a metrically stressed syllable at their onset. Analysis of a large, phonemically transcribed corpus by Cutler and Carter (1987) showed that 1 in 3 English content words start with a stressed syllable. Furthermore, since these items are more frequent (mostly because monosyllabic words, which are all stress-initial, are of high token frequency), a strategy of placing word boundaries before stressed

---

[1] Associate Professor of linguistics in Department of English language, Imam Khomeini International University of Qazvin (corresponding author); vsadeghi@hum.ikiu.ac.ir
[2] MA Graduated of Linguistics, Imam Khomeini International University of Qazvin; mohammadinegin65@gmail.com

syllables would correctly locate the onsets of 90% of content words in the London-Lund corpus of spoken conversation.

Experimental evidence also suggests that the presence of a strong syllable is used by listeners as a cue to the start of a new word. The word-spotting paradigm has been used to show that listeners are faster to detect monosyllables followed by a strong syllable than by an unstressed (weak) syllable (Cutler & Norris, 1988; Norris, McQueen, & Cutler, 1995; Vroomen, van Zon, & de Gelder, 1996; see McQueen, 1996, for a review of research using the word spotting task).

The MSS has also been proposed as an account of how infants learn to divide the speech stream into words. It has been demonstrated that English-speaking 9-month-old infants display a preference for hearing words that conform to the predominant strong-weak stress pattern (Echols, Crowhurst, & Childers, 1997; Jusczyk, Cutler, & Redanz, 1993; Morgan, 1996; see Jusczyk, 1997 for a review of this and related work).

The MSS as described will operate successfully for both open-class that begins with an unstressed syllable and content words that begin with a stressed syllable (Cutler & Carter, 1987). For closed-class words, the reverse (weak-initial) stress pattern is generally found. In the corpus investigated by Cutler and Carter (1987), 69% of weak syllables are at the onsets of closed-class words, with fewer than 5% being the initial syllables of open-class words. In order to utilize the MSS effectively Cutler and Carter propose that two separate strategies operate for accessing the lexical representations of words in separate stores of open- and closed-class items. Strong initial syllables are used to access the open-class lexicon while words beginning with unstressed syllables are looked up in the store of closed-class words. Evidence supporting this dual-lexicon and dual-access strategy account comes from naturally occurring 'slips-of-the-ear' and laboratory-induced boundary misperceptions (Cutler & Butterfield, 1992). Listeners are more likely to incorrectly add a word boundary before a strong syllable and are more likely to delete word boundaries before weak syllables. In adding or removing words from an utterance, these misperceptions tended to preserve the relationship between initial stress and lexical class, words created from weak syllables were more likely to be closed-class and words with strong initial syllables more likely to be open-class.

2. Materials and method

In the present research, we conducted a perception experiment to determine to what extent native speakers of Persian are able to identify meaningful words from nonsense words (non-words). A number of non-words were created by adding a syllable to the beginning or end of some Persian existing bisyllabic words. The trisyllablic non-words were then produced by a native speaker in experimental condition. In the next step, the non-word spoken stimuli were played for a number of Persian speakers. Participants were asked to identify the meaningful words in the stimuli. Response time (RT) was measured from the offset of the stimulus to the time of the key press. The collected RTs were log-transformed for statistical analysis.

3. Results and discussion

Results suggested that Persian listeners identified meaningful words from stress-final non-words more reliably and faster than stress-initial or stress-medial non-words. In addition, results indicated that the more stress placement in non-words

agrees with patterns of stress distribution in Persian words, the faster and more reliable the process of spoken word recognition, while the less agreement between stress placement in non-words and patterns of Persian words' stress distribution exists, the slower and less reliable the process of spoken word recognition can be.

4. Conclusion

Overall, the results of this study show that Persian listeners use patterns of prosodic variation to identify word boundaries when linguistic (such as syntactic or semantic) information is not available in the spoken stimuli. This strongly confirms the Metrical Segmentation Hypothesis, suggesting that stress affects spoken word recognition in Persian.

Keywords: Metrical segmentation hypothesis; Non-words; Speech recognition; Stress; Word boundaries