CHAPTER TWELVE

SYLLABLES IN SPEECH PROCESSING: EVIDENCE FROM PERCEPTUAL EPENTHESIS

Andries W. Coetzee

1 Syllables in Speech Production and Processing

Much of phonological theory assumes that speech is both discrete and hierarchical. The speech stream can be divided into discrete units (segments/sounds), and these units are hierarchically organized into larger chunks such as syllables, feet, prosodic words, etc. (Nespor and Vogel 1986, Selkirk 1978, 1986, etc.). Phonological rules are often formalized in terms of these higher order groupings of sounds. This approach presents many problems to phonological theory. Not only is it difficult to define these higher order categories accurately, but it is probably even harder to find evidence of their reality.

Speech is transferred between speaker and listener as continuously varying air pressure fluctuations. The structure of the pressure fluctuation is sometimes such that discontinuities can be detected easily—for instance at the transition between a plosive and a vowel. This lends some support to the idea that speech is discrete—the speech stream can be divided into discrete chunks based on its physical properties. However, the boundaries between segments are not always demarcated clearly in the physical speech stream, as anyone who has ever tried to find the boundary between a vowel and glide on a spectrogram or waveform can attest. The fact that it is not always possible to find correlates to concepts like segments in the physical world has lead some to doubt the existence of discrete segments (Port and Leary 2005).

Just as difficult as finding physical evidence for the division of the speech stream into discrete segments, is finding physical evidence for the grouping of these segments into larger prosodically defined units such as feet, syllables, etc. Although there are temporal and intonational properties of the speech stream that have been shown to correlate with the boundaries of some of these larger prosodic units (Beckman and Edwards 1990, Byrd and Saltzman 2003, Coetzee and Wissing 2007, Selkirk 1981, 2001, Sugahara 2005, Turk and Shattuck-Hufnagel 2000,
etc.), this correlation is not very strong, as is clear from chapter 8 of this Handbook. The observed variation in these temporal and intonational measures is often large, and the differences between different prosodic units are often very small (McQueen 1998, 21).

Given the difficulty of finding consistent physical correlates for these higher prosodic units, a question that needs answering is whether (all of) these units are real. If physical evidence of these units is lacking, is there evidence that language users parse the speech stream in terms of these units? If a positive answer can be given to this question, then it can be concluded that some of these prosodic units are at least psychologically real—that is, even if they do not exist in the physical properties of the speech stream, the mind of the language user imposes these structures onto the speech stream in the process of planning speech production and perceiving speech. These structures may then have mental reality even if they lack physical reality.

This chapter discusses evidence that listeners do parse the speech stream into higher order prosodic groupings during the process of speech perception. Specifically, it shows that listeners are sensitive to syllabically defined allophonic distributions. In English, aspirated voiceless stops appear only in syllable initial position. When presented with an acoustic signal such as [spʰika] with an unambiguous aspirated [pʰ], listeners impose a percept on this stimulus where [pʰ] appears in syllable initial position. This leaves the [s] prosodically stranded—there is no vowel in the actual acoustic stimulus to which the [s] can affiliate. The evidence presented below shows that the listeners “perceive” a vowel between the [s] and [pʰ] in this kind of situation, i.e. they perceive [sәp hika]. The [ә] that forms part of the percept is not present in the physical stimulus, and is therefore imposed onto the percept by the perceptual system of the listener. This is evidence that the perceptual system does impose higher order structure onto the speech stream, and that this part of the speech processing process is even robust enough to result in percepts for which there is no physical evidence.

Before getting into the details of the current study, we briefly review some existing evidence for the mental reality of such higher order prosodic structures in both production and perception. It is not the goal of this section to review the extensive literature about this, but rather to discuss just a few examples of the evidence that language users do use syllables both in production and perception. Similar evidence for other prosodic constituents also exists (Clifton et al. 2006, Schafer et al. 2000, Wightman et al. 1992, etc.).