Journal of Memory and Language 64 (2011) 460–476

Contents lists available at ScienceDirect

Journal of Memory and Language

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

Cues and cue interactions in segmenting words in fluent speech

Rochelle S. Newman [a,*], James R. Sawusch [b], Tyler Wunnenberg [c]

[a] Dept. of Hearing & Speech Sciences, 0100 Lefrak Hall, University of Maryland, College Park, MD 20742, United States
[b] Dept. of Psychology, State University of New York at Buffalo, 206 Park Hall, Buffalo, NY 14260-4110, United States
[c] University of Iowa, United States

ARTICLE INFO

Article history:
Received 19 July 2009
revision received 13 November 2010
Available online 4 March 2011

Keywords:
Segmentation
Lexical access
Word recognition

ABSTRACT

Fluent speech does not contain obvious breaks to word boundaries, yet there are a number of cues that listeners can use to help them segment the speech stream. Most of these cues have been investigated in isolation from one another. In previous work, Norris, McQueen, Cutler, and Butterfield (1997) suggested that listeners use a Possible Word Constraint when segmenting fluent speech into individual words. This constraint limits the word recognition system to consider only those parsings that could conceivably be words in the language (that is, those that do not strand illegal sequences). The present paper examines how this constraint interacts with other cues to segmentation, such as junctural and allophonic cues and neighborhood probabilities. Segmentation was influenced both by the PWC and by the presence of acoustic cues to juncture, such as the acoustic results of a speaker's intention to produce a particular phoneme as the end of one syllable vs. as the start of another (vuff-apple vs. vuh-fapple). In contrast, segmentation was not affected by the legality of a syllable-final vowel (tense vs. lax), or by the similarity of a sequence to words. This suggests that acoustic cues in the signal play a far larger role in segmentation than do sources of bias from the lexicon, and that probabilistic lexical information from the lexicon (such as neighborhood information) is unlikely to be used in the process of word segmentation.

Introduction

When we hear a person speaking, we have the illusion that we are listening to a series of individual words, one following another in an orderly procession. Yet what actually hits our ear is an ever-varying pattern of air pressure changes, without any clear breaks indicating where one word ends and another begins (Cole & Jakimik, 1980; Klatt, 1980; Reddy, 1976). This is particularly problematic because many long English words contain shorter words embedded within them (McQueen, Cutler, Briscoe, & Norris, 1995). Listeners cannot simply wait until they hear the end of a word and be assured that the next sound is the beginning of a subsequent word. Understanding how listeners break the steady speech stream into its component words remains one of the fundamental issues in speech perception and word recognition research.

Most current theories assume that word recognition takes place through the simultaneous satisfaction of multiple constraints (see McQueen (2005) for a review). These constraints may include acoustic–phonetic, phonological, lexical, prosodic, syntactic, semantic, and contextual information. According to this approach, many possible words may be considered simultaneously, and the set of possibilities evolves through the course of processing as more possibilities, and more constraints, come into play. Attempts at connectionist modeling have shown that networks relying on multiple constraints or strategies perform much better than those limited to a single cue (Christiansen, Allen, & Seidenberg, 1998). However, until recently, most of the behavioral research on these constraints has examined them individually, in isolation from one another. How different cues interact during the course of language

* Corresponding author. Fax: +1 301 314 2023.
E-mail address: rnewman@hesp.umd.edu (R.S. Newman).

0749-596X/$ - see front matter © 2011 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2010.11.004

processing has received less attention (but see Mattys, 2004; Mattys, White, & Melhorn, 2005).

Among the sources of information that have been examined as cues to word segmentation in English are the allophonic details of how phonemes are realized in different syllable positions (Davis, Marslen-Wilson, & Gaskell, 2002; Nakatani & O'Connor-Dukes, 1979), phonotactic probabilities (Gaygen & Luce, 2002; McQueen, 1998), stress and metrical information (Cutler & Norris, 1988; Mattys, 2004; Mattys et al., 2005), and the Possible Word Constraint (PWC, Norris, McQueen, Cutler, Butterfield, & Kearns, 2001; Norris et al., 1997). Different theoretical approaches focus to a greater or lesser extent on subsets of these cues (Mattys et al., 2005). For example, some theories focus more heavily on cues to word-boundary locations such as phonotactic probability, stress, and junctural and allophonic cues (e.g. Shortlist, Norris, 1994). Other theories focus more on the task of identifying the words themselves, rather than their boundaries (the neighborhood activation model, Luce & Pisoni, 1998) even as they exploit detailed acoustic–phonetic information related to phoneme position and syllable/word boundaries. Boundaries are thus a result of an interaction among cues in the signal and competition among alternative word candidates. This is the case in some connectionist models of word recognition such as TRACE (McClelland & Elman, 1986). Despite these different theoretical perspectives, most research has focused on testing the effects of individual cues, rather than exploring how these cues may be used in concert.

Mattys et al. have offered a theoretical framework for how these cues interact in perception as a hierarchy of constraints. At the top level is lexical knowledge with semantics, pragmatics and knowing which sequences of sounds make words in the language (lexical knowledge) as the most important sources of information. At the next tier, segmental information related to phonemes constrains segmentation. This information reflects influences of pronunciation and coarticulation, language-specific details of syllable structure, and phonotactic constraints. Finally, at the lowest tier is information related to lexical stress. In English this would include syllable duration, amplitude and the voice pitch profile as the acoustic correlates of stress. In support of this hierarchy Mattys (2004; Mattys et al., 2005) has presented data from multiple experimental tasks that show that when information at a higher level in the hierarchy is available to listeners, the influence of information at a lower level is reduced or absent in listener performance. For example, with clear speech, phonetic/allophonic information influenced segmentation but stress did not. When the speech signal was degraded with noise, phonetic/allophonic information did not influence segmentation but stress information did (Mattys, 2004). Mattys et al. refer to this intermediate tier of information as segmental cues. For clarity, however, we will refer to it as phonetic/allophonic to avoid confusion with the use of the word "segmentation" to describe the process of dividing the continuous speech signal into words.

The present work seeks to complement that of Mattys et al. by further exploring which cues may be more important for segmentation in situations in which multiple cues are present simultaneously. First, it is possible that the various sources of information within a level of the Mattys et al. framework are not all equally effective as cues to segmentation. In particular, we focus on junctural/allophonic and vowel identity (a probabilistic phonotactic cue), both of which occur at the phonetic/allophonic level in Mattys' hierarchy, to determine whether such cues have different weightings in speech segmentation.

It is also unclear whether certain types of information belong at one level or another within this hierarchy. We explore two cues that were not included in Mattys' work: the Possible Word Constraint (PWC) of Norris et al. (1997) and lexical neighborhood effects. In the PWC, segmentation is guided by a constraint where each "unit" that results from the segmentation process must be a potential/possible word in the language. Norris et al. (2001) have formulated this as a language universal constraint where the syllable is the smallest unit that can be a word and all syllables contain a vowel. Thus, when a listener is asked to recognize words in spoken nonsense utterances, the word *apple* will be easier to spot in *vuffapple* than in *fapple*. This is because the remaining *vuff* in *vuffapple* is a well-formed syllable and could be a word in English. The isolated *f* in *fapple* cannot be an English word. Results consistent with this have been reported for English (Norris et al., 1997), Japanese (McQueen, Otake, & Cutler, 2001), and Sesotho (Cutler, Demuth, & McQueen, 2002). On the one hand, the PWC may be presumed to be part of the lexical information level, as it involves a restriction on what items can be a word in the language, and thus results from the knowledge of words rather than the phonetic/allophonic tier of information. However, it also has a great deal in common with phonotactic constraints (as per Gaygen & Luce, 2002), which are part of the phonetic/allophonic information level of the Mattys et al. (2005) hierarchy.

The present experiments thus investigate the PWC, the allophonic details of how phonemes are produced in syllable-initial and -final position (juncture cues), the role of vowel identity in syllable structure (phoneme probabilities within and across syllable/word boundaries), and the similarity of a nonsense syllable to real words as sources of word segmentation information. Following Norris et al. (1997), we will use the word-spotting task where on some trials a real word is preceded by other phonemes that results in a nonsense utterance and the task for listeners is to indicate whether each utterance contained an embedded word or not. Before proceeding with the individual experiments, we will provide a brief review of a variety of cues used in segmentation.

Allophonics/juncture cues

In fluent speech (and the word-spotting task) the acoustic details of how phonetic sequences are pronounced varies with syllable position. In English, the detailed acoustic realization of a phoneme is often different when it begins a stressed syllable than when it occurs in other syllable positions or occurs at the beginning of an unstressed syllable (Lehiste, 1960; Umeda & Coker, 1974). For example, the phonemic sequence /gretaI/ can be pronounced as "grey tie" or as "great eye". In "grey tie", the /e/ vowel at the

end of "grey" is relatively long and the syllable initial /t/ is aspirated with a relatively long-VOT. In "great eye", the /e/ in "great" is relatively short, the /t/ is not aspirated and has a short VOT, and the onset of the vowel /aI/ ("eye") is often glottalized or laryngealized. Nakatani and Dukes (1977) and Umeda and Coker (1974) showed that talkers of American English systematically produce different variants (allophones) of the voiceless stops, the approximant /l/, and vowels for different syllable positions. Moreover, listeners are sensitive to these subtle acoustic differences, both in natural speech (Nakatani & O'Connor-Dukes, 1979) and in synthetic speech (Dutton, 1992; Repp, Liberman, Eccardt, & Pesetsky, 1978), and can use them to help locate word boundaries during online processing (Davis et al., 2002; Shatzman & McQueen, 2006a, 2006b). We will refer to these differences in how a phoneme is realized, depending upon the position of the phoneme in the syllable, as juncture cues. Church (1987) proposed that listeners use this information to segment the speech signal into words.

This junctural information (along with phonotactic information) is typically used perceptually during syllabification, and several studies have suggested that syllabification is an important component of segmentation and word recognition (Content, Dumay, & Frauenfelder, 2000; Dumay, Frauenfelder, & Content, 2001). We choose not to use the term syllabification because this term does not distinguish between the division into syllables based on acoustic/phonetic vs. phonotactic properties, which we separate in the present paper. Moreover, focusing on the acoustic junctural properties in particular allows us to predict that the ease of segmentation may depend on the particular consonant occurring at the boundary and the extent to which it shows position-based allophonic variation.

Probabilistic phonotactics

Many languages have constraints on what sounds can begin or end a syllable or word. For example, in English, content words generally do not end with a lax vowel. Thus, listeners may use the type of vowel in a syllable as a source of information for subsequent segmentation and avoid placing a segmentation boundary immediately after a lax vowel. In the original PWC study (Norris et al., 1997), context syllables always contained a lax vowel. That is, the authors chose to consistently use items such as *vuffapple* rather than items such as *veefapple*. The lax vowel may have influenced listeners to place a segmentation boundary after the consonant rather than before it, making segmentation easier. This alternative is similar to the metrical segmentation strategy proposed by Cutler and Norris (1988). Norris et al. (2001) investigated this possibility explicitly, and found no overall effect of vowel quality, although they did find some weak trends when the target words were strong–weak (a strong or stressed syllable followed by a weak or unstressed syllable). As the target items were not cross-spliced in Norris et al., there remains the possibility that other aspects of pronunciation differed across vowel conditions. Still, their results suggest that vowel quality is unlikely to have a substantial effect on segmentation. We decided to explore this cue despite these results because vowel quality could still interact with other cues to segmentation, even if it is not a sufficiently strong cue to affect segmentation by itself.

Neighborhoods/similarity to words

Finally, a number of studies of word recognition have documented that the similarity of a target word to other real words in the mental lexicon (the neighborhood) can influence both word recognition (e.g. Luce & Large, 2001; Vitevitch, 2002) and phoneme perception (Newman, Sawusch, & Luce, 1997, 2005). It is unclear whether the neighborhood of the "nonsense" syllable that would remain when a listener segments a target word from a nonsense sequence will influence performance. As this information is based on lexical identity, it is generated from a higher level in the Mattys et al. (2005) hierarchy. A fuller discussion of this source of information and its possible role in segmentation is contained in Experiment 2, below.

Experiment 1 investigated the combined influence of the Possible Word Constraint, junctural and allophonic cues, and probabilistic phonotactics (vowels in syllable-final position) in a word-spotting task to determine the relative roles of these constraints during segmentation. This will allow us to examine the role of each source of information during the process of segmentation. This experiment also provides a further test of the Possible Word Constraint itself. If the driving force in Norris et al.'s effect was the constraint that isolated consonants cannot be a word in English (all English words must contain a vowel), listeners should perform better on sequences resulting in possible rather than impossible words, even when other cues to segmentation (such as junctural cues) are controlled and varied independently. Listeners should be able to spot the word in items such as *vee-fapple* more easily (faster and/or more accurately) than in *fapple*. Experiment 1 was designed to extend the results of previous studies by manipulating several different cues to segmentation at the phonetic/allophonic tier orthogonally. This allows us to examine how these cues jointly influence the process of segmentation. Within the hierarchical framework of Mattys et al. (2005), we are examining whether the PWC, allophonic details, and probabilistic phonotactics will all influence listener performance since they all represent phonetic/phonological information.

Experiment 2 examined a type of knowledge-based or lexical cue, lexical neighborhood. If word segmentation is treated as a constraint satisfaction system, then the probability that a sequence is a word should influence segmentation. In turn, this would imply that the more word-like the precursor portion of the carrier was, the easier it would be for listeners to spot the embedded word. One factor that makes an item "word-like" is its similarity to other words in the language, or its lexical neighborhood. Thus, precursor syllables with dense lexical neighborhoods (such as /fɪp/) might make it easier to spot a following word than precursor syllables with few lexical neighbors (such as /zɛp/).

Unlike the acoustic–phonetic cues described above, the information necessary for lexical neighborhood effects is based on information stored in memory, rather than from

the signal, *per se*. Moreover, research suggests that competition among lexical candidates should only occur after the process of segmentation leads to the competitor set (Vroomen & de Gelder, 1995; Norris, McQueen, & Cutler, 1995). However, such competition may still affect segmentation decisions that occur at subsequent points in the speech stream. These issues were explored in Experiment 2.

Experiment 1

The first experiment was designed to extend the results of previous Possible Word Constraint studies by manipulating several different cues to segmentation orthogonally. In addition to the effect of the Possible Word Constraint, we examined three further factors that might play a role in word segmentation: two related to allophonic details, and (third) the phonotactic constraint that words not end in a lax vowel.

Juncture cues

The stimuli here were recorded so that in half the cases, the speaker intended to produce the sequence with the syllabic boundary before the boundary consonant (i.e., *vuh-fapple*) and in the other cases the speaker intended to produce the sequence with the syllabic boundary after the boundary consonant (i.e., *vuff-apple*). Consistent with prior studies, we predict that listeners will find it easier to hear the word when the consonant is part of the prior syllable than when it is attached to the word itself.

It is important to note, here, the role of coarticulation in the allophonic details of phonemes. The production of a phoneme is influenced by the preceding and following phonemes. Furthermore, the phenomenon of coarticulation occurs across syllable boundaries (Öhman, 1966) and listeners are sensitive to coarticulatory information across syllable boundaries (Martin & Bunnell, 1982). The degree of coarticulation is moderated, however, by the prosodic structure of an utterance (Fougeron & Keating, 1997). To the extent that productions are more extreme and less reduced for consonants at the beginning of a stressed (vs. unstressed) syllable, then the allophones that signal juncture cues can be thought of as resulting from a reduction or change in coarticulation.

Consonant class

Phonemes with consistent, robust allophonic cues include the voiceless stop consonants /p/, /t/, and /k/, and the liquid /l/ (Christie, 1974; Nakatani & Dukes, 1977; Umeda & Coker, 1974). Previous research has identified reliable acoustic correlates to phoneme position for these phonemes. In contrast, Lehiste (1960) found that the only potential cues for distinguishing syllable-initial from syllable-final fricatives were the durations of the preceding vowel, the duration of the fricative, and the location of the amplitude minimum. These acoustic qualities appeared to be less robust (smaller and less consistent). In the present experiment, half of the items contained boundary consonants that typically have robust (strong) allophonic differences while the other half contained boundary consonants with less reliable (weak) allophonic differences. We predict that this will interact with production constraints. The speakers' intention to produce a syllable-final vs. syllable-initial consonant will have a larger effect on those phonemes with clear allophonic variants (voiceless stops and /l/) and these variants will have a larger influence on listener's ability to spot words. For ease of discourse, we will refer to this as a difference in consonant class (or consonant for short).

Probabilistic phonotactics

In English, syllables generally do not end in a lax vowel. In fact, it is sometimes claimed that English syllables cannot end with a lax vowel, though interjections such as "eh" and "huh" indicate that the prohibition is limited to lexical items. Norris et al. (2001) describe this as a constraint on content words. With the exception of function words, vowels such as /ɛ/ and /ɪ/ are almost always followed by a consonant within the same syllable. In comparison, English syllables frequently end with tense vowels such as the /i/ at the end of the first syllable in "sequence" or the /o/ in "motion". Half of the items in the present study had tense vowels in the first syllable and the other half had lax vowels that usually require a following consonant. If the phoneme-to-phoneme probabilities within syllables in English are an important source of constraint in speech segmentation, then lax vowels will tend to pull the following consonant towards them perceptually. Indeed, studies on syllabification suggest that intervocalic consonants tend to "be linked with, or 'stick' to, a lax, stressed vowel" (Derwing, 1992, p. 224–225; see also Treiman & Danis, 1988). Listeners should, therefore, find it easier to hear the word when the vowel in the preceding syllable is lax (*vuff-apple*) than when it is tense (*veefapple*). This effect may interact with consonant class and/or junctural cues. Effects of the vowel may be weaker or nonexistent when the consonant was produced as part of the first syllable (in which case it follows the vowel regardless of vowel identity), and the effect may be stronger for consonants with weaker allophonic constraints (where the syllabification of the consonant is less clear-cut). It is also possible that lax vowels would simply cause consonants to be treated as ambisyllabic, instead of syllable-initial, and such a change might have no implications at all for the subsequent syllable.

These cues (phonotactics, strength of allophonic variation or consonant class, and junctural cues) are typically correlated in speech. The present experiment varies these cues orthogonally in order to determine the circumstances under which each of these cues operates in the word-spotting (segmentation) task. While it would be interesting to orthogonally combine these sources of information with the PWC, this is not possible with natural, fluent speech and all of the other cues that we have chosen to study. First, since the point of the PWC is that the /f/ (an isolated consonant) can not form a word by itself, there can be no phoneme prior to the /f/. This means that the quality of the prior vowel can not be orthogonally combined with the PWC. Secondly, while fricatives, such as /f/, can be produced in isolation and it might be possible for a talker to produce *fapple* with both a syllable initial /f/ and a syllable final /f/, a talker cannot produce *poven* with a syllable final

/p/. Stops are produced as a movement from one articulatory configuration (in this case, silence or rest) to another (the initial vowel of "oven"). This typically results in a long-VOT, syllable-initial /p/. Attempting to splice the syllable-final /p/ from another utterance onto the word "oven" would likely result in acoustic correlates of the phoneme prior to the /p/ (the prior vowel) being included and would sound unnatural. Since the intent was to study how listeners segment fluent speech, the PWC could not be orthogonally combined with vowel quality and pronunciation intent. It can be (and was) combined orthogonally with consonant class.

We used as target words the 48 words used by Norris et al. (1997), all of which began with a vowel and contained stress on the first syllable. To these we added an additional 12, bringing the total number to 60 (see Appendix A in supplementary material). Of these words, 30 were monosyllabic (e.g., EARTH), and the other 30 were bisyllabic (APPLE). We then created five versions of each word: one with a single phoneme precursor (/l/EARTH, /f/APPLE), and four with full-syllable precursors. These four syllable precursors represented the orthogonal combination of vowel quality and pronunciation intent (juncture cues). One had a lax vowel and was pronounced with a syllable-initial consonant (/rɛ-l/EARTH, /v∧-f/APPLE), one had a lax vowel but was pronounced with a syllable-final consonant (/rɛl/-EARTH, /v∧f/-APPLE), one had a tense vowel and a syllable-initial consonant (/yo-l/EARTH, /vi-f/APPLE) and the fourth had a tense vowel and a syllable-final consonant (/yol/-EARTH, /vif/-APPLE). Across all five versions, half of the words were preceded by consonants with strong allophonic cues (/l/, /p/, /t/, or /k/), and half were preceded by consonants with weak allophonic cues (/v/, /f/, /ʃ/, /θ/, /z/, /tʃ/, /s/, or /ʤ/) (see Lehiste, 1960).

In order to ensure that pronunciation of the target word varied only when intended and that the impossible condition (e.g., /l/EARTH) did not differ in pronunciation from the possible items, we cross-spliced across items with the same pronunciation intent (same second syllable). Thus, /rɛ-l/EARTH, /l/EARTH, and /yo-l/EARTH all had the identical second syllable (and embedded word), ensuring that the pronunciation of the word and of the boundary consonant were identical in all three cases. We also cross-spliced between /rɛl/-EARTH and /yol/-EARTH, such that the word was pronounced identically in these two cases as well. This method of cross-splicing maintained the pronunciation difference between those items that were intentionally produced differently (/yol/-EARTH vs. /yo-l/EARTH; /rɛ-l/EARTH vs. /rɛl/-EARTH), while removing other pronunciation differences in the target words. That said, it is worth noting that the cross-splices also result in the modification of some differences that would occur in natural speech, such as the syllable shortening that occurs in longer words – in typical speech, the word "earth" should be shorter in /yo-l/EARTH than in lEARTH. To avoid having this be a confound, we balanced which word served as the base for cross-splicing across items. Thus, if the target word INK was taken from a natural production of /vI-t/INK, the target word AGE might be taken from the natural production of tAGE.

Finally, we measured the acoustic details of our talker's actual pronunciations and report the results of these measurements. This allows us to compare our data to previous measurements and assess the degree to which the items contained the intended junctural cues.

Method

Listeners

Fifty-six students at the University of Iowa participated in this experiment in exchange for course credit. All were native speakers of English and had no history of a speech or hearing disorder. Each listener heard each target word only once. Listeners were divided into five groups with members of each group hearing a different version of each word. Data from 14 participants were excluded from analysis for the following reasons: one fell asleep, one was left-handed (this results in slower and more variable reaction times to the critical word stimuli), and four responded on less than 85% of the trials. Five had very low overall accuracy scores (less than 60% correct) and three had high false alarm rates to the nonwords. This left a total of 42 participants overall.

Stimuli

The 48 words used by Norris et al. (1997) were supplemented with an additional 12 words following the same constraints, as described above. Each word was produced once when preceded by a single consonant, and four times when preceded by a full syllable. These syllables consisted of the crossing of two factors: pronunciation intent (CV-Cword vs. CVC-word) and vowel type (lax vs. tense). Thus, each word occurred in five different conditions. For ease of discussion, we will call these conditions the *fapple*, *vuff-apple*, *vuh-fapple*, *veef-apple*, and *vee-fapple* conditions. Appendix A in supplementary material shows the complete set of items.

For half of the target words, the boundary consonant was selected to have strong allophonic cues to syllable position: these consisted of the phonemes /p/, /t/, /k/, and /l/. For the other half of the words, the boundary consonants were fricatives or affricates, chosen because they have much weaker allophonic cues (these include the sounds typically written as "sh", "ch", "j", "f", "v", "th", "z", and "s"). The stimuli of Norris et al. (1997) all used fricatives, affricates and nasals for their boundary consonants.

Listeners heard one version of each target word, yielding 12 words of each type (*fapple*, *vuff-apple*, *vuh-fapple*, *veef-apple*, and *vee-fapple*) for each listener. This means that listeners only heard 12 of the 60 target items in the single-consonant + word condition (compared to 48 in the syllable + word conditions). In order to prevent strategic effects, we recorded an additional 36 filler items consisting of a single consonant followed by a vowel-initial word. In half of these items the word was a single syllable (such as /g/EACH, which contains the word each) and in the remainder the embedded word was bisyllabic (such as /d/ACID, containing the word acid). Thus, listeners heard a total of 96 items with embedded words; 48 of these had a single consonant as the precursor portion, and the other 48 had a full CVC syllable as the precursor portion. Half

of the words for each were monosyllabic and half were bisyllabic. This resulted in a set of carrier items in which 25% of the items were monosyllabic (those consisting of a single consonant precursor with a monosyllabic word), 50% were bisyllabic (those consisting of either a single consonant precursor with a bisyllabic word or a syllable precursor with a monosyllabic word) and 25% were trisyllabic (those consisting of a syllable precursor and a bisyllabic word).

We also created 96 distracter items that did not contain embedded words. These had the same syllabic breakdown as the items containing embedded words: 24 were monosyllabic, 48 were bisyllabic, and 24 were trisyllabic. Finally, we created a set of 14 practice items of varying lengths (one, two, or three syllables) of which six contained real words.

All items were recorded by a female native speaker of English at a 44.1 kHz sampling rate with 16 bits quantization and stored on computer disk. We then proceeded to cross-splice the target syllables of the items such that matched sets of items contained the same embedded word or consonant + embedded word. For items in the *vuff-apple* and *veef-apple* conditions, the second (and third) syllables from one of the two target items were replaced with the identical portion from the other target item, resulting in two items with different first syllables but equivalent embedded words. For half of these items, the original recording with a lax vowel (e.g., *vuff-apple*) was used as the base, and for the other half the original recording with a tense vowel (e.g., *veef-apple*) was used as the base. For items in the *fapple*, *vuh-fapple* and *vee-fapple* conditions, the syllable or syllables beginning with the boundary consonant from one of the three recordings served as the base and replaced the identical portions from the other two items. Thus, the "fapple" portion of all three of these items was identical. In one third of the cases the original *fapple* recording served as the base, and in one third of the cases each of the other two items was used as the base.

These final target items were then divided into five sets of items, each containing one version of each word, and an equal number of each type of word. Thus, each condition contained 12 items in the *fapple* condition and 12 each in *vuff-apple*, *vuh-fapple*, *veef-apple*, and *vee-fapple* conditions. Each set also included all 36 filler items and all 96 distracters, for a total of 192 items.

Procedure

Each listener heard a 14-item practice block, followed by all 192 items of his or her condition in a single test block. Listeners were instructed to press the far left button on their response box any time they heard an item that did not contain an embedded word, and to press the far right button on their response box any time they heard an item that did contain an embedded word. Both speed and accuracy were emphasized.

We did not ask listeners to respond to the items out loud. This represents a change in methodology from Norris et al. (1997). In Norris et al., whenever a listener indicated the presence of a word target, they also pronounced the word aloud. This allowed the researchers to check that listeners had heard the intended word and had not false-alarmed to an alternative word. In this experiment, we examined individuals' responding to the nonwords, and, as noted above, removed from analysis any individual with a high rate of false alarms. On average, the rate of identifying words in the target items was 50% greater than the rate of identifying words in the nonword items (that is, of false alarming), suggesting that these remaining participants were quite accurate. However, to ensure that false alarm rates were not affecting our results, we reran all subjects analyses including only those individuals who had at least 85% accuracy on the nonword items. Differences between these analyses are noted in the text.

Results

Acoustic measurements

One concern is that the instructions to the speaker were either ineffective, or caused the speaker to produce items in an unnatural manner. In order to ensure that pronunciation intent did result in the expected acoustic differences between items, we performed an acoustic analysis of the productions (see Table 1). The acoustic correlates that were chosen for measurement are ones that have been identified as differing on the basis of syllable position in previous studies (particularly Lehiste, 1960; but see also Dutton, 1992; Hoard, 1966; Umeda & Coker, 1974, 1975). By comparing productions intended as having a syllable-initial consonant (-CV) to those intended as having a syllable-final one (C–V), we can assess whether our speaker's intentions influenced the acoustics of these syllables. We discuss the strong-allophonic consonants first, followed by the weak-allophonic consonants.

For the voiceless stop consonants /p, t, k/, we examined the voice onset time (VOT), the duration of the stop closure, and the voicing duration of segment(s) preceding the stop. (The latter is intended as a correlate to the duration of the preceding vowel. However, since vowels could not always be separated from preceding consonants, we measured the duration of the entire voiced portion, not merely the vowel itself.) VOT has long been viewed as the primary cue differentiating word-initial and word-final voiceless stops (Umeda & Coker, 1974, 1975). We expect that stop consonants produced so as to be word-initial will have longer VOTs than those stops intended as word-final. Some speakers appear to lengthen their closure duration in CV contexts, although this is variable (Dukes & Nakatani, 1976). Both vowels and nasals tend to be shorter when followed by a stop consonant in the same syllable than when the stop consonant occurs at the onset of the following syllable (Dukes & Nakatani, 1976; but see Quené, 1992).

For the /l/, we measured F_1 and F_2 at the point of maximum F_3/minimum amplitude. Dukes and Nakatani (1976) found that the second formant of /l/ consistently reached a lower frequency in word-final position than in word-initial position. Using synthetic speech, Dutton (1992) found that F_1 had an even greater effect on listeners' perception of /l/ than F_2, so we measured this as well. A 14-pole LPC analysis with a 256-sample Hamming window was used for the formant analyses, after downsampling to an 11,025 Hz sampling rate. Peaks in the

Table 1
Acoustic measurements of consonants in syllable-initial and -final position. Durations are in ms and frequencies in Hz. Values in parentheses are standard errors.

Voiceless stops /p, t, k/	Vowel duration	Closure duration	VOT duration	Laryngealization (count max 50)
Initial (-CV)	136 (7.9)	65 (1.9)	65 (1.8)	0
Final (C–V)	137 (7.3)	69 (2.3)	21 (1.5)	42
Approximant /l/	F_1 frequency	F_2 frequency		Laryngealization (count max 10)
Initial (-CV)	429 (11.9)	1346 (21.2)	0	
Final (C–V)	513 (24.6)	1133 (25.0)	6	
Affricates	Vowel duration	Closure duration	Consonant duration (not incl. closure)	Laryngealization (count max 10)
Initial (-CV)	161 (13.6)	56 (4.8)	98 (7.8)	0
Final (C–V)	168 (13.4)	43 (3.1)	69 (3.9)	10
Fricatives	Vowel duration		Consonant duration	Laryngealization (count max 50)
Initial (-CV)	144 (7.8)	110 (5.1)	2	
Final (C–V)	154 (7.1)	95 (2.8)	41	

LPC were identified as formants and checked against a wide-band spectrogram.

For both stops and /l/, the occurrence of laryngealization of the following vowel was tabulated. Laryngealization is manifested by irregularly spaced, low amplitude glottal pulses and often occurs before word-initial vowels (Dilley, Shattuck-Hufnagel, & Ostendorf, 1996; Dukes & Nakatani, 1976). Laryngealization has been shown to influence perception in several studies (Dutton, 1992; Nakatani & Dukes, 1977), and appears not to be strongly affected by the nature of the preceding consonant (Dilley et al., 1996).

For the voiceless stops, no consistent difference in preceding vowel duration was found ($t(49) = -0.18$), perhaps as a result of variation in speaking rate across stimuli. There was a small, non-significant difference in closure duration in the expected direction, with longer closure when the stop consonant was at the end of the first syllable, $t(49) = 1.52$, $p > .10$. There was a large and consistent difference in VOT, $t(49) = 22.3$, $p < .001$, with longer VOTs in the syllable-initial stops. All of the C–V VOTs were longer than all of the C–V VOTs, with only a single exception. Both the ranges of VOTs and the difference between CV and C–V VOTs are similar to those previously reported for voiceless stops (see Lehiste, 1960; Umeda & Coker, 1974, 1975). For the /l/, F_1 values were lower in syllable-initial position, $t(9) = -2.87$, $p < .05$, and F_2 values were higher in syllable-initial position, $t(9) = 5.09$, $p < .001$, as expected. There was also a difference in laryngealization in the stop consonant and /l/ syllables. None of the CV stimuli showed signs of being laryngealized (having a lower amplitude and lower F_0) while 48 out of 60 of the C–V items were laryngealized, a significant difference ($\chi^2 = 240$, $p < .001$). Interestingly, the three consistent cues found here (VOT for stops, F_1 for /l/, and laryngealization) were also the cues shown by Dutton (1992) to be most important in perceptual studies for distinguishing different sequences.

Consonants with weak allophonic cues included both fricatives and affricates. The measurements were analogous to those for the voiceless stop consonants: preceding vocalic duration, closure duration for affricates, frication duration, and the presence of laryngealization. A number of researchers have reported that consonants tend to be longer in syllable-initial position (Hoard, 1966; Klatt, 1976; Lehiste, 1960) and that this duration difference influences listeners' perception (Christie, 1977; Quené, 1991, 1992; Shatzman & McQueen, 2006a, 2006b).

For the affricates, we again found no difference in preceding vowel duration ($t(9) = -0.66$), but did find significant differences in closure duration ($t(9) = 2.87$, $p < .05$) and frication duration ($t(9) = 5.72$, $p < .001$) in the expected direction (see Table 1). In measuring the duration of the affricates and fricatives, the presence of aperiodicity in the waveform (frication) was used to assess duration. For the fricatives we found a small but consistent difference in preceding vowel duration, $t(49) = -2.02$, $p < .05$. However, this is in the opposite direction from that expected on the basis of prior research (Dukes & Nakatani, 1976; Lehiste, 1960). There was a small but consistent difference in the duration of the fricative, $t(49) = 3.44$, $p < .005$, with the syllable-initial fricatives longer in duration. There was also a difference in laryngealization in the fricative and affricate syllables: only two of the CV stimuli showed signs of being laryngealized (having a lower amplitude and lower F_0), whereas 51 out of 60 C–V items contained the acoustic correlates of laryngealization. This difference was also significant ($\chi^2 = 313$, $p < .001$).

These results show that the speaker succeeded in altering the pronunciation of these items in ways consistent with a change in syllabification and prior research. In particular, consonants were lengthened in word-initial position, and laryngealization occurred when syllables began with a vowel. In both the strong and weak allophonic cases there were clear acoustic qualities that could have been used by the listener as cues to syllabification and word boundaries.

Perceptual results

Response times were measured from stimulus offset, as in Norris et al. (1997). Since the embedded word targets differ in their duration and the task essentially requires that listeners hear all of the target word before responding to it, the end of the target appears to be the most appropriate reference point for RTs. Any response time greater than

two standard deviations from the condition mean for that listener was excluded from the analysis.[1]

In the analysis of the data, two ANOVAs were run: one with subjects as the random factor (F_1) and one with items as the random factor (F_2). In most cases, these two analyses give the same results and we make special mention of any analysis in which they do not. Five items had accuracy scores of 0, and thus had no reaction times. The average reaction time for the category was used as the RT measure for these items in the items analysis. We also provide partial η^2 values on the subjects analyses as measures of effect size. These represent the proportion of variance in the dependent variable explained or accounted for by the differences in the means for the effect hypothesis tested. Partial η^2 is the variance attributable to the effect divided by that of the effect and error (Cohen, 1988). Given the large number of independent variables, this is more appropriate than using the total variance as the denominator.

The results can be seen in Fig. 1 for the accuracy data, and Fig. 2 for the reaction time data. In general, these figures show substantially larger error rates, and slower reaction times, for the CV than the C–V items, and that strong allophonic cues result in slower reaction times and greater error rates in the C–V items, but in lower error rates among the C–V items. The data also show that violating the PWC and leaving an isolated consonant behind (*fapple*) produced the longest overall RTs. Finally, there is no apparent influence of vowel quality (probabilistic phonotactics) in the data. These patterns are explored more fully in the statistical analyses below.

The results were examined with two main analysis designs. In the first analysis, we explored the five word types (*fappe*, *vee-fapple*, *vuh-fapple*, *veef-apple*, and *vuff-apple*) and the effect of consonant class in a 2×5 ANOVA. In the second analysis, we excluded the *fapple* condition. This allowed us to reanalyze the data in a $2 \times 2 \times 2$ design, exploring the effects of juncture cues, consonant class and probabilistic phonotactics simultaneously. We describe each of these two approaches below, along with the specific predictions that we made for each one. In order to simplify the flow of the text, the text itself describes the general patterns and the actual statistical results from the ANOVAs are given in Table 2.

The first analysis focused on the factors of consonant class (consonants with strong vs. weak allophonic cues) and word type (*fapple*, *vee-fapple*, *vuh-fapple*, *veef-apple*, and *vuff-apple*). This approach examines the Possible Word Constraint, and whether it varies with consonant class. It

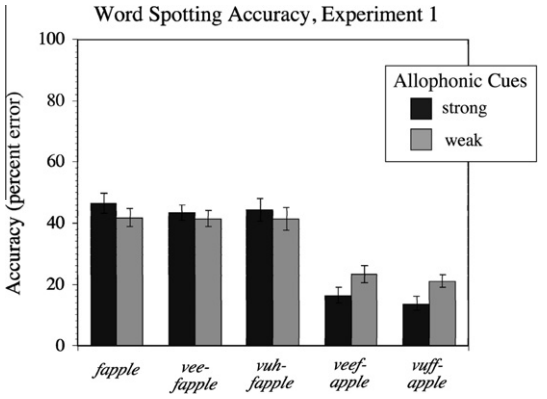

Fig. 1. Percentage error in word-monitoring, on the basis of the identity of the precursor syllable. Error bars reflect standard error.

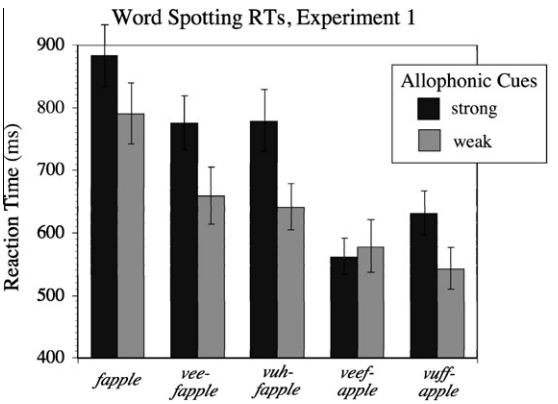

Fig. 2. Reaction times in word-monitoring, on the basis of the identity of the precursor syllable. Error bars reflect standard error.

also allows us to conduct follow-up analyses comparing specific word types to one another. In general, we would expect, based on the PWC, that the embedded word would be harder to spot in items such as *vee-fapple* and *fapple*, where identification of the word "apple" leaves "f" as a remainder, than in items such as *veef-apple*. As a result, the main effect of word type should be significant, which it was (see Table 2 for specific analysis results).

If the consonant class also has an effect on segmentation, we would expect to find that the effect of word type would interact with that of consonant. We predict an interaction, rather than a main effect, because the presence of strong allophonic cues should make it easier to spot the word in conditions such as *veef-apple* and *vuff-apple*, but harder to spot the word in conditions such as *vee-fapple* and *vuh-fapple*, where the allophonic cues serve to make the consonant appear to be part of the target word. In fact, the accuracy data showed no main effect of consonant, but did show a significant word type by consonant interaction in the subject analysis only. The reaction time data, in contrast, showed a significant main effect of consonant but no interaction, again in the subjects analysis only. These effects involving the consonant are the only situation in which subject and item analyses differed, a fact that will be discussed later. Assuming for the moment that the

[1] Norris et al. (1997) treated trials with long RTs as errors. Their method has the disadvantage that if a listener is particularly slow in one condition, this would also show up as poorer accuracy, as many of their slow responses would be considered wrong. In order to ensure that the method of analysis did not influence our results, we actually did all analyses three ways. In addition to removing trials with RTs greater than two standard deviations from the condition mean for the listener, we performed the analysis as Norris et al. did, with any RT > 1750 ms being counted as an error. Third, we removed all response times greater than 3000 ms from the analysis, and then used harmonic means of the RTs in the statistical analyses. Ratcliff (1993) has shown that the harmonic mean is much less sensitive to spurious, long RTs (outliers) than the arithmetic mean. There were few differences among these three methods and none would alter the interpretation of the data. Consequently, we report only the one analysis.

Table 2
Statistical results, Experiment 1.

Effect	Accuracy			Reaction times		
	By subjects	By items	Partial η^2	By subjects	By items	Partial η^2
2 × 5 ANOVA						
Word type	$F_1(4164) = 51.3$, $p < .0001$	$F_2(4232) = 29.7$, $p < .0001$	0.238	$F_1(4164) = 29.1$, $p < .0001$	$F_2(4232) = 17.2$, $p < .0001$	0.415
Consonant class	$F_1 < 1$	$F_2 < 1$		$F_1(1, 41) = 16.7$, $p < .0001$	$F_2 < 1$	0.290
Word type by consonant interaction	$F_1(4164) = 2.58$, $p < .05$	$F_2(4232) = 1.3$, $p > .10$	0.059	$F_1(4164) = 1.73$, $p > .10$	$F_2 < 1$	
2 × 2 × 2 ANOVA						
Pronunciation	$F_1(1, 41) = 183.3$, $p < .0001$	$F_2(1, 58) = 63.0$, $p < .0001$	0.8175	$F_1(1, 41) = 67.8$, $p < .0001$	$F_2(1, 58) = 22.8$, $p < .0001$	0.985
Vowel	$F_1 < 1$	$F_2 < 1$		$F_1 < 1$	$F_2 < 1$	
Consonant class	$F_1(1, 41) = 3.67$ $p < .01$	$F_2 < 1$	0.082	$F_1(1, 41) = 12.1$, $p < .001$	$F_2 < 1$	0.923
Pronunciation × vowel	$F_1 < 1$	$F_2 < 1$		$F_1 < 1$	$F_2 < 1$	
Pronunciation × consonant	$F_1(1, 41) = 6.56$, $p < .01$	$F_2(1, 58) = 2.16$, $p > .05$	0.138	$F_1(1, 41) = 4.67$, $p < .05$;	$F_2(1, 58) = 1.87$, $p > .10$	0.711
Vowel × consonant	$F_1 < 1$	$F_2 < 1$		$F_1(1, 41) = 2.46$, $p > .05$	$F_2 < 1$	
3-Way interaction	$F_1 < 1$	$F_2 < 1$		$F_1 < 1$	$F_2(1, 58) = 1.57$, $p > .10$	

effects of consonant are real, two observations should be noted. First, for both the *vuhf-apple* and *veef-apple* conditions (taken together), the consonants with strong junctural cues led to greater accuracy, although in the former case there appears to be a speed–accuracy trade-off, making interpretation more difficult (see the right side of Figs. 1 and 2). In contrast, for the *fapple*, *vee-fapple*, and *vuh-fapple* conditions where the listener had to ignore the junctural cues and re-parse the signal to identify the word, the strong junctural cue consonants produced both slower and more error-prone responses (see Table 3), although only the reaction time effects reached statistical significance. Thus, the influence of consonant class (strong vs. weak junctural cues) is modulated by the pronunciation intent of the talker and whether it is consistent with the embedded word.

Although the main effect of word type is expected based on the PWC, the PWC makes a more specific prediction: spotting the word in conditions such as *fapple* should be more difficult than those such as *veef-apple*. However, there are actually two possible interpretations of the PWC. One possibility is that junctural cues determine the location of syllable boundaries prior to any effect of the PWC. If so, "apple" is misaligned with the syllable boundary in both *vee-fapple* and *fapple*. The word should be harder to spot in both of these items than in a sequence such as *veef-apple*. This appears to be the prediction made by Norris et al. (1997, 2001), and is in fact the case (for *vee-fapple* vs. *veef-apple*, $t_1(41) = 8.36$, $t_2(59) = 6.37$, both $p < .0001$, $\eta^2 = 0.630$ by accuracy and $t_1(41) = 5.39$, $p < .0001$, $t_2(59) = 3.21$, $p < .0005$ by RTs, $\eta^2 = 0.415$; for *fapple* vs *veef-apple*, $t_1(41) = 8.04$, $t_2(59) = 6.55$, $\eta^2 = 0.612$ by accuracy and $t_1(41) = 8.99$, $t_2(59) = 7.24$ by RTs, $\eta^2 = 0.663$ all $p < .0001$).[2] However, such a result could be explained by the junctural cues themselves, without reference to the existence of a constraint against possible words. The acoustic cues to juncture suggest that "apple" is its own word in one case, while "fapple" is a word in the other case. It is therefore not surprising that hearing the word would be easier in the former condition.

A stronger test of the PWC is to compare the *fapple* condition with the *vee-fapple* condition. Here, neither junctural cues nor vowel quality cues predict a difference in word spotting. Only a constraint against stranding single consonants differs between the two cases. In the accuracy data, there was no significant difference between the *fapple* and *vee-fapple* conditions (both $t < 1$). However, there was such an effect in the reaction time data ($t_1(41) = 3.28$, $p < .005$, $t_2(59) = 3.00$, $p < .005$, $\eta^2 = 0.208$). Listeners were slower in the *fapple* condition, averaging 837 ms, and faster in the *vee-fapple* condition where they averaged 712 ms (left side of Fig. 2). This replicates and confirms the Possible Word Constraint findings across variation in pronunciation. The fact that such a difference was not found in the accuracy data, as it had been in Norris et al. (1997), suggests that some of their comparison items may have differed both in terms of the PWC and in terms of the junctural cues present in the

[2] These analyses ignore the distinction between items with a weak junctural cue at the boundary and those with a strong junctural cue at the boundary. Thus, "fapple" includes items like "fapple" but also like "kedge".

Table 3
The effect of consonant class (consonants with strong vs. weak allophonic cues to juncture) in each of the five word types.

Item	Measure	Strong junctural cues	Weak junctural cues	*t*-Test
vuhf-apple	Reaction times	631 ms	543 ms	$t(41) = 2.61$, $p < .05$
	Accuracy	86.3%	79.0%	$t(41) = 2.80$, $p < .01$
veef-apple	Reaction times	562 ms	578 ms	$t(41) = -0.42$, $p > .05$
	Accuracy	83.7%	76.6%	$t(41) = 2.23$, $p < .05$
fapple	Reaction times	884 ms	790 ms	$t(41) = 1.70$, $p < .10$
	Accuracy	55.5%	58.3%	$t(41) = -1.40$, $p > .05$
vuh-fapple	Reaction times	779 ms	641 ms	$t(41) = 2.54$, $p < .05$
	Accuracy	55.7%	58.7%	$t(41) = -0.66$, $p > .05$
vee-fapple	Reaction times	776 ms	659 ms	$t(41) = 2.88$, $p < .01$
	Accuracy	56.5%	58.5%	$t(41) = -0.59$, $p > .05$

stimuli. In our study, while the *fapple, vuh-fapple,* and *vee-fapple* conditions all resulted in accuracy levels between 50% and 55%, the *vuff-apple* and *veef-apple* conditions showed over 15% higher accuracy levels.

It is also possible to view both the *fapple* and *vee-fapple* conditions as "stranding" the /f/; in *fapple*, the preceding silence serves to strand the /f/, whereas in the *vee-fapple* case, the junctural cues do so. According to this view, the difference between the *fapple* and *vee-fapple* conditions is not so much a test of the Possible Word Constraint, but instead a test of the strength of different segmentation cues; the significant difference suggests that there is a greater penalty for stranding the /f/ when the cues for the word boundary preceding it (silence) are stronger. Regardless of interpretation, the general pattern of the present results appears to support the presence of a constraint on possible words.

The second analysis allows us to explore the consonant effects from the first analysis in greater depth, as well as exploring how the different potential segmentation cues interact more generally. Setting aside the *fapple* condition, we are left with a $2 \times 2 \times 2$ design representing all combinations of pronunciation (juncture cues in CV-Cword vs. CVC-word), vowel quality (a probabilistic phonotactic constraint on ending syllables with a tense vs. lax vowel) and consonant class (stronger allophonic differences vs. weaker allophonic differences).

A three-way ANOVA was used to explore this subset of the data. Pronunciation had a substantial effect on listeners' accuracy and reaction times, with listeners more accurate and faster at detecting the embedded word when the syllable boundary coincided with the word boundary. Listeners made 18% errors and responded in only 579 ms when the word boundary and syllable boundary coincided (as in *veef-apple* and *vuff-apple*), but made 42% errors and responded in 714 ms when they did not (*vee-fapple* and *vuh-fapple*).

Vowel did not produce a significant main effect nor did it result in any significant interactions in either the accuracy or RT data. Apparently, listeners did not reliably exploit the probability that a syllable can end in a tense vs. lax vowel when making their embedded word responses (in line with prior results of Norris et al., 2001). This indicates that listeners' segmentation in this task was based primarily on acoustic information in the signal, an issue to be explored further in Experiment 2.

Finally, the consonant class had a significant effect both on listeners' accuracy and reaction times in the subject analysis only, with slower reaction times to items with strong allophonic cues at the word boundary than to items with weaker allophonic variation. The pronunciation by consonant interaction was likewise significant in the subject analysis only, in both the accuracy and reaction time data. In the RT data, the pronunciation by consonant interaction reflects longer RTs for the strong juncture cue consonants when the boundary consonant is syllable initial and conflicts with segmenting at the target word onset. In the accuracy data, the interaction reflects higher accuracy (lower error rates) for the strong cues when the pronunciation is consistent with segmentation of the target word. As noted earlier, some of the influence in the accuracy data may reflect a speed/accuracy trade-off. Overall, conflict between strong juncture cues and word segmentation hurts performance while agreement between the strong juncture cues and the word boundary facilitated performance relative to the influence of weak juncture cues.

Both in this analysis, and the prior analysis, effects involving the consonant are the only situation in which the items analysis and subjects analysis differ. This raises the question as to whether these consonant effects are real or generalizeable effects. Since the number of items per condition was relatively small, it is possible that variability among the items is simply swamping a real effect, and the large effect sizes support this interpretation. Another possible explanation for this discrepancy is that it is caused by the particular groupings of consonants. The analysis compares consonants with "stronger allophonic cues" to those with "weaker allophonic cues" – but both of these two sets include a variety of consonants. The different phonemes in the stronger allophonic cue condition may not all be equivalent in the degree to which they contain junctural cues. However, based on the acoustic analysis, the most likely reason is that the set of consonants in the weaker allophonic variation group actually contain strong junctural cues, at least for some items. A consistent difference was found for this set in the laryngealization of the word/syllable-initial vowel, and previous perceptual studies (see Dutton, 1992) have shown that listeners are very sensitive to this acoustic cue. The grouping into two sets is clearly an over-simplification, and thus finding null effects in these item analyses is not terribly surprising.

In order to explore this more fully, we also examined the reaction time measures for the individual items, to see if any patterns could be identified. First, the general pattern (of slower RTs for the items with "strong allophonic cues",

particularly when the consonant was syllable initial), seemed to hold. That is, it did not seem to be generated by a small subset of the particular items. Nor did there appear to be any items that were atypical among the items in the strong allophonic condition. However, there were seven items in the weak allophonic condition that did appear to be atypical. These items had exceptionally long RTs, particularly in the consonant-initial conditions. These included three items where /v/ was the medial consonant, two involving /f/, and one each involving /ʃ/ and /θ/. It appears that the lack of significant effects in the item analysis is the result of variability among the consonants in the weak allophonic condition, such that some of these consonants influenced segmentation in ways similar to and overlapping with the strong juncture cue consonants. As noted in the acoustic analyses, there were systematic differences in the acoustic correlates to consonant identity for these consonants. Thus, it is not surprising if listeners were able to exploit this information, particularly for some items. This increases the variability in the items analysis and results in the non-significant differences. Experiments involving individual consonants (e.g. Dutton, 1992; Shatzman & McQueen, 2006a, 2006b) are needed to explore in more depth the role of acoustic correlates to phoneme position for individual sequences of phonemes.

Experiment 2

The results of Experiment 1 suggest that listeners use a variety of constraints during the process of segmentation. The Possible Word Constraint is one such constraint, which serves to bias against sequences that cannot be words in the language. Listeners also use real words as a constraint. White, Melhorn, & Mattys (2010) found stronger priming for "corridor" following "anythingcorri" than after "imoshingcorri" – the existing real word caused listeners to posit a boundary before "corr" in the first case, but not the second.

In addition to the distinction between items that are possible words and those that are impossible, sequences may also vary in the likelihood (or probability) of their being a word. That is, there may be gradations in word-likeness among sequences, and this, too, could influence segmentation judgments. One way of measuring such probabilities is by lexical neighborhood (Luce & Pisoni, 1998). Cluff and Luce (1990) showed that the neighborhood of the embedded words that make up spondees (words that consist of two equally-stressed monosyllables, such as baseball) influences listener accuracy in identifying spondees presented against a noise background. Studies have also shown that the perception of nonsense syllables is affected by the number and frequency of occurrence of similar-sounding words in the language (Newman et al., 1997, 2005). Items that are similar to more real words in the language appear to be treated as if they were themselves more "word-like". Since listeners seem to parse sequences in ways that result in possible words, they may also be more likely to parse sequences in ways that result in more probable words, or in nonwords that are more similar to a greater number of words. The current experiment investigates this prediction by comparing listeners' detection of words in sequences where the syllable before the target word comes from a high density lexical neighborhood or a low density neighborhood.

However, other work suggests that lexical biases, particularly those based on cohort size, may only have an effect after the segmentation process has been completed. For example, Vroomen and de Gelder (1995) examined priming for onset-embedded words when the words were followed by syllables of differing cohorts. Their results demonstrated that cohort size mattered only when there was overlap between the syllable in question and the target embedded word. Thus, Dutch listeners were slower to detect "MELK" in *melkaam* than in *melkeum*, presumably because many more Dutch words begin with the sequence *kaa-* than with *keu-*. Since *kaa-* is a more frequent word onset than *keu-*, it activates many more lexical entries, which then compete with the activation of *melk*. However, cohort size had no effect in their study when the potential sequences did not overlap. Listeners were no slower to detect "BEL" in *belkaam* than in *belkeum*. This pattern of results suggests that while the probability of something being a word can influence the extent of competition among different interpretations, it does not influence the likelihood of segmentation itself. Similarly, Norris et al. (1995) used a word-spotting task in which words occurred at the onset of a two-syllable string (e.g., "minteff"). They found an interaction between cohort size (e.g., the number of words that begin with the 2nd medial consonant) and second-syllable stress: when the second syllable included a strong vowel (initiating segmentation), the number of members in the cohort had an influence on speed of identifying the first word. But when the second syllable was unstressed, there was no competitor effect. Thus, competition effects occurred only after segmentation based on metrical structure. Likewise, Cutler and Butterfield, in a study exploring missegmentations of faint speech, found that errors were not higher in frequency than the intended input, suggesting that segmentation was not driven by lexical competition.

While these results suggest that lexical neighborhood will not directly affect segmentation, we felt that a direct test was still warranted. These papers imply that competition occurs only after segmentation has identified a potential word onset. However, most sentences will contain multiple potential word onsets that are identified sequentially during the course of the sentence. As a result, even if competition among potential lexical items occurs only after segmentation has occurred at that syllable location, this would not preclude those competition effects from influencing segmentation of subsequent syllables. That is, even if the neighborhood properties of a second syllable cannot influence segmentation preceding it, the neighborhood properties of a first syllable could potentially affect segmentation following it. This is, in essence, the argument made by van der Lugt (2001), that the position of phonological information relevant to segmentation (at the offset preceding the target or the onset of the target) can alter the influence of phonotactic constraints, and presumably other constraints as well. In both the Norris et al. and Vroomen and de Gelder studies, the target word occurred at the onset of the two-word sequence, with potential competition effects occurring on subsequent syllables. In the present

study, the effect of lexical competition occurs on a first syllable, and thus could potentially influence segmentation of the subsequent syllable.

In addition, Vroomen and de Gelder (1995) and Norris et al. (1995) used cohort size (words that begin with the same initial sounds) as their measure of similarity to words. Experiment 2 uses lexical neighborhood as its measure. While these two measures are often correlated, they are separable. Similarity to real words based on both measures seems to influence word recognition (Allopenna, Magnuson, & Tanenhaus, 1998; Magnuson, Dixon, Tanenhaus, & Aslin, 2007; Newman et al., 2005; Vitevitch, 2002). Consequently, despite the prior work by Vroomen and de Gelder and by Norris et al., we felt that a test of the influence of neighborhood of the stranded syllable on segmentation of a subsequent word was justified.

Since all of the items in this experiment contained juncture cues to segmentation and the items were all presented in the clear (no noise), information at the phonetic/allophonic tier of the Mattys et al. (2005) hierarchy should be exploited in the word-spotting task (see Mattys, 2004). If similarity to real words is not a part of the phonetic/allophonic tier of information, or it is simply not an effective segmentation cue when preceding a word boundary, then the prediction is clear. Our results should mirror Vroomen and de Gelder and there should be no effect of the similarity of the preceding phoneme sequence to words (neighborhood) on word spotting. Alternatively, similarity of the preceeding phoneme sequence to real words (neighborhood) may act as one of the sources of constraint at the phonetic/allophonic tier that is exploited by listeners to segment speech.

Method

Subjects

Fifty members of the University of Iowa community participated in this experiment in exchange for course credit. All were native speakers of English, and had no history of a speech or hearing disorder. Data from an additional six participants were excluded from analysis for the following reasons: two for being non-native speakers, one for equipment failure, and three for experimenter error. As before, we ran the analysis twice, once including all of the other participants, and the other time including only those who both responded on a minimum of 80% of the trials, and had accuracy of at least 85% on the nonword items. The pattern of results remained the same across the two analyses.

As each listener could hear a target word only once, the listeners were divided into two groups of 25 listeners each, with members of each group hearing a different version of each word.

Stimuli

We used the same words and boundary consonants as in the previous experiment. We created two different nonsense-word syllable precursors for each boundary consonant: one syllable had a high neighborhood value and was similar to more, and more common, words. The second syllable had a low neighborhood value with fewer and less common similar words. The complete set of items is shown in Appendix B in supplementary material. Although neighborhood calculations were based on the CVC (such as /wʊk/ in /wʊ-k/EDGE), the initial CV portions also differed in neighborhood between the two sets. Thus, regardless of a listener's segmentation, one set results in a remainder that is similar to more real words than does the other.

Neighborhood calculations were performed in the same manner as that of Newman et al. (1997). Each CVC nonword was compared to real words in an on-line dictionary. We defined an item's neighbors as being every real word that differed from the target by only a single phoneme substitution, addition, or deletion. After determining the number of neighbors for each item, neighbors were weighted by their log-transformed frequencies and summed to yield a frequency-weighted neighborhood density (henceforth, neighborhood). Thus, words that are more common in the language contribute more to our calculations of lexical neighborhood than do rare words. Only those neighbors with a familiarity index of 6.0 or greater on a 7-point rating scale (Nusbaum, Pisoni, & Davis, 1984) were included to avoid basing our calculations on words unlikely to be in our listeners' lexicons.

Half of the items had a voiceless stop consonant or an /l/ as the boundary consonant (a consonant with strong allophonic cues to syllable position), and the other half had fricatives or affricates (with weaker allophonic cues), as in Experiment 1. However, stop consonants tend to be much more frequent word-finally in English than are fricatives. Thus, it was not possible to match the size of the neighborhood differences across these different consonant sets. For that reason, the analyses in this experiment collapse across these different boundary consonants.

All of the items were produced as CV-Cword, rather than CVC-word. This should encourage the final consonant of the precursor syllable to bind with the word in question. This allows us to examine whether differences in neighborhood make it easier (or more difficult) to separate the second consonant from the target word.

We also used 60 of the distracter items from Experiment 1, and we created a set of 12 practice items. All of the items were recorded by a male native speaker of English at a 44.1 kHz sampling rate, with 16 bit quantization, edited, and stored on computer disk. We then proceeded to cross-splice the target syllables of the items, such that matched pairs of items contained the same embedded word + boundary consonant.

These final target items were then divided into 2 sets of items, each containing one version of each word. Each set contained half of the words in the high-neighborhood condition and half in the low-neighborhood condition.

Procedure

All of the listeners heard a 12-item practice block, followed by all 120 items of their set in a single test block. The listeners were instructed to press the far left button on their response box any time they heard an item that did not contain an embedded word, and to press the far right button on their response box any time they heard

an item that did contain an embedded word. Both speed and accuracy were emphasized.

Results and discussion

Response times were measured from the offset of the stimuli, as in Norris et al. (1997) and Experiment 1. As in Experiment 1, any response time greater than two standard deviations from the condition mean for a listener was eliminated from the analysis. The mean accuracy and reaction times for the high and low density conditions are shown in Fig. 3.

We first examined listeners' error rates. While listeners were slightly more accurate at detecting words in the high-neighborhood condition, this effect was not significant (high neighborhood, 34.9% error rate; low neighborhood, 38.0% error rate; $t_1(49) = 1.59$, $p > .10$; $t_2(59) = 0.64$, $p > .10$). The effect in the reaction times was significant by subjects only ($t_1(49) = 2.42$, $p < .05$; $t_2(59) = 0.02$, $p > .10$)[3] and was in the opposite direction of the trend in the accuracy results (high neighborhood response times averaged 626 ms, low neighborhoods averaged 600 ms).[4] That is, listeners were both slower, and more accurate, in the high-neighborhood condition. This would seem to indicate a speed–accuracy trade-off, but these effects were only significant by subjects. Differences in neighborhood may have encouraged listeners to adopt different response strategies, but we did not find any reliable evidence that they affected listeners' segmentation.

These results are consistent with and extend the prior results of Vroomen and de Gelder (1995) and Norris et al. (1995). Across two languages (Dutch and American English), two ways of measuring the similarity of remaining segments to words (cohort and neighborhood), and two experimental tasks (priming and word spotting), a consistent pattern of results has been found. The similarity of the remaining segments to real words does not seem to affect segmentation, at least when reliable junctural cues to segmentation are available. Moreover, this appears to be the case regardless of whether the lexical competition is occurring prior to the potential segmentation boundary or subsequent to it.

General discussion

These two experiments examined a number of potential cues to word segmentation, including junctural cues, consonant class (strength of junctural cues), and probabilistic phonotactics (probability that a syllable ends with a vowel) as well as both the possibility and probability of a remaining sequence of phonemes being a word in the language. The segmentation cues that were directly based on

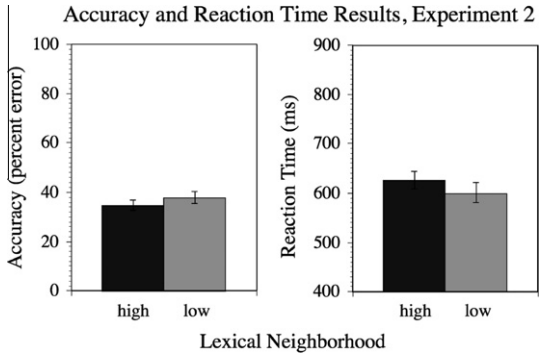

Fig. 3. Percentage error and reaction times for word-monitoring in Experiment 2, on the basis of the neighborhood of the precursor phrase.

acoustic correlates in the speech signal all had an influence on listeners' speed and/or accuracy in spotting words in the nonsense strings.

First, variations in production that result in junctural cues in the allophonic realization of phonemes appear to play a large role in word segmentation. Listeners found it substantially easier to find the embedded word when a sequence was produced with a syllable boundary coinciding with the word boundary (*vuff-apple*, rather than *vuh-fapple*). This effect was somewhat stronger for consonants with strong allophonic cues to syllable position (such as stop consonants), but was also present for those consonants with weaker allophonic variations. Based on prior research and the acoustic measurements of the stimuli used here, laryngealization of the vowel at the beginning of a syllable/word seems to be a robust cue to word and syllable boundaries, and listeners seem to exploit this cue in segmenting speech into words. We also found evidence that talkers produce subtle differences in frication duration for fricatives and affricates in different syllable positions. However, because of the presence of a correlated acoustic cue, laryngealization of the vowel at word onset, further research will be needed to explore the ability of listeners to exploit the acoustic cues to consonant position in fricatives and affricates (see Shatzman & McQueen, 2006a, 2006b for data on the fricative /s/ followed by /p/ or /t/).

Experiment 1 also replicated results of Norris et al. (1997) that it was easier for listeners to find embedded words when the remainder of the sequence could form a legal word in the language than when it could not form such a word. This was the case even when other cues that often correlate with the Possible Word Constraint, such as junctural cues, were varied independently. Within the Mattys et al. (2005) hierarchy of cues to segmentation, this implies that the Possible Word Constraint is a part of the phonetic/allophonic tier of information since the presence of one does not preclude the influence of the other when both vary.

In contrast, the probabilistic phonotactics of a syllable-final vowel did not appear to be taken into consideration by listeners, at least not in the experiments here. Despite the fact that lax vowels are rare in syllable-final position in English content words and tense vowels are more frequent at the end a syllable, listeners did not find it easier (or harder) to spot embedded words across these two types

[3] One item had 0% accuracy in one condition, and thus had no reaction time measure. The mean reaction time of the category was used as the value for this item.

[4] We re-examined the results by replacing items with RTs or accuracy more than two standard deviations away from the mean of the category with the category mean. This did not change the RT results, but the accuracy results in the subjects analysis did become significant, $t(49) = 2.30$, $p < .05$. Again, however, the high neighborhood items were both slower and more accurate, making interpretation difficult.

of vowels. This result replicates Norris et al. (2001) using a more controlled set of stimuli. The present results go beyond those of Norris et al. in manipulating vowel identity orthogonally with pronunciation and consonant class. The complete lack of any effect of vowel identity here, coupled with the results of Norris et al. thus indicates that segmentation in English, across both American and British versions, does not exploit the phonotactic constraint of vowel identity (tense vs. lax) when there are clear junctural cues present. This result is also consistent with those of van der Lugt (2001) who found that phonotactic cues at syllable offset did not influence segmentation.

There are also some results in the literature that may be seen as contradicting this conclusion. First, several studies on syllabification have reported that listeners tend to consider a medial consonant as part of the first syllable when the vowel is lax (Derwing, 1992; Treiman & Danis, 1988). Those studies used listeners' judgments of which production of real-words sounded more natural (i.e., "meh-lon" vs. "mel-un"). The present studies suggest that this bias against syllable-final lax vowels may be post-perceptual, and thus not influence the process of segmentation. Second, Kirk (2001) reported that listeners found it easier to detect the word "lunch" when the preceding string was /vit^h/ rather than /vik^h/. In English, /l/ cannot follow an aspirated /t/at the beginning of a syllable, and the fact that listeners found it easier to detect the word following such a phoneme suggests that they were using phonotactic restrictions to guide segmentation. One reason for the difference may be the strength of the phonotactic information. In the present study, such information was generally probabilistic in nature, as compared to involving language-based prohibitions, such as in Kirk. That is, sequences in our study may have been unlikely words, but not impossible. A second difference is that Kirk's phonotactic manipulation was at word onset while ours was at word offset (see van der Lugt, 2001).

Differences in neighborhood for a remainder also did not appear to influence segmentation. Listeners did not appear to find it easier (or harder) to identify an embedded word when the initial sequence was, itself, more likely to be a word in English than when it was less likely (cf. Vroomen & de Gelder, 1995). There was, however, some evidence that this probability difference may have affected listeners' response strategies. Overall, the results of Experiments 1 and 2 show that probabilistic knowledge-based constraints driven by the identity of sequences of phonemes (whether a syllable can end in a lax vowel) and by similarity of the stranded syllable to words in the mental lexicon do not influence segmentation of a clear speech signal into words when junctural cues are present. Again, using the hierarchy of cues to segmentation of Mattys et al. (2005), this implies that the probabilistic phonotactic information and neighborhood information in the present experiments are not a part of the phonetic/allophonic tier of information. Either these cues are not exploited by listeners at all or they are only exploited when the allophonic details (junctural cues) are obscured, such as in a noisy speech signal.

One concern is whether these results may be specific to the word-spotting task, a task that is admittedly somewhat unnatural. We think that this is unlikely. Listeners consistently exploited the acoustic correlates to phoneme position, the allophonic details of the phoneme, in this word-spotting task. These results replicate earlier studies of speech acoustics and perception of word sequences using both natural and synthetic speech (e.g., Dutton, 1992; Nakatani & Dukes, 1977). This similarity in the effects of allophonic details indicates that their use, by listeners, is similar across the tasks. In turn, this suggests that listeners would exploit the PWC in normal speech (speech with only real words) in a fashion similar to that found using the word-spotting task. Finally, data from infants also show that the exploitation of juncture cues has priority over statistical regularities (Johnson & Jusczyk, 2001).

Another concern is whether the present results are limited to words beginning with vowels. Work with infants has suggested that words beginning with vowels are harder to segment than the more common, consonant-initial words (Jusczyk & Aslin, 1995; Mattys & Jusczyk, 2001), and this may be the case for adult listeners as well. However, it seems unlikely that listeners would rely on different cues when segmenting vowel-initial words, simply because they would not normally know the word would begin with a vowel unless they had already successfully segmented it. Thus, it would seem likely that the pattern of results found here would generalize to other types of items. This conclusion is consistent with recent studies by Shatzman and McQueen (2006a, 2006b) on segmentation with /sp/ and /sk/ sequences (where the words could start with /s/ or the stop) in Dutch that show that listeners are sensitive to the allophonic details of phoneme position.

Norris et al. (2001) describe the Possible Word Constraint as based on the linguistic universal that all syllables contain a vowel. Since words are, in turn, built from one or more syllables, all words must contain a vowel. The word "apple" is difficult to segment out of the nonsense sequence "fapple" because the remaining /f/ left behind does not include a vowel and thus cannot be a word. This would seem to indicate that listeners should always find it harder to segment a word when the sequence left behind is a consonant than when it is a vowel. Recent results for Slovak (Hanulikova, McQueen, & Mitterer, 2010), where an isolated consonant can be a word, show that there is no segmentation penalty for stranding an isolated consonant when it is a meaningful unit. Thus, the PWC is not a universal constraint against isolated consonants. In addition, for Japanese listeners, the moraic consonant /n/ is as easy to segment from a word as a vowel (McQueen et al., 2001). One possibility for explaining this is that there are acoustic correlates to the position of a moraic consonant in Japanese. This would make segmenting a moraic /n/ in Japanese similar to segmentation in English where there are strong acoustic correlates to phoneme position and syllable boundary. In this case, the results of McQueen et al. for moraic /n/ do not reflect the operation of the PWC *per se*.

An alternative is that since the rhythmic structure of Japanese seems to be based on moraic units (see Han, 1994; Port, Dalby, & O'Dell, 1987), the prohibition in Japanese is against a strand that is not a rhythmic unit. In English, syllables are the carrier of rhythmic structure.

This means that a syllable can be left as a remainder and an isolated consonant cannot. This raises the interesting question of the nature of the acoustic information that indicates whether the remainder can or can not be a word. For a speaker of English, is it that the strand must include a vowel (any vowel), as it appears that Norris et al. (2001) propose? Alternatively, are there acoustic correlates to the rhythmic units of speech, perhaps based on fundamental frequency and amplitude envelope, that contribute directly to segmentation without mediation by the phonetic class of the segment? If this is the case, it would seem to contradict the hierarchy of segmentation cues proposed by Mattys et al. (2005) since acoustic correlates of the rhythmic structure of speech should be a part of the stress tier yet the PWC appears to operate at the phonetic/allophonic tier. Finally, as we noted earlier, is it simply the strength and number of allophonic details that also drives the operation of the Possible Word Constraint, at least for English and Japanese? Resolution of this question must be left to future research.

The current findings have a number of implications for our understanding of the architecture of the lexical segmentation system. First, as noted above, although listeners used the PWC in the present experiments, they did not use it to the exclusion of other cues. That is, the presence of this cue did not supersede entirely other potential cues for segmentation. As such, it does not appear to function in the same way as does lexical status in the Mattys et al. (2005) study. This may suggest it is better considered to be operating at the acoustic/phonetic level of the hierarchy than at the lexical level.

Second, at least in the present task, probabilistic cues at the acoustic-segmental level seemed to outweigh probabilistic cues based on lexical information (such as neighborhoods and probabilistic phonotactics). Prior work suggested that lexical competition occurs only after segmentation (or, that the act of segmentation provides the basis for the set of lexical items to be activated), and does not influence segmentation. Yet one might nonetheless have expected that competition among entries early in a speech string would still have an influence on segmentation of subsequent boundaries. Surprisingly, this was not the case. It may be that the lexical level of segmentation is limited entirely to information about legality, rather than information that is probabilistic in nature. That is, the act of finding word boundaries may be based on a process of satisfying multiple constraints *from the signal*, but this process does not appear to include constraints from other levels of processing, even when processing at those other levels would have already begun. Information on what *is* a word in the language appears to behave quite differently from information on what is *likely* to be a word in the language.

Finally, even among sources of information putatively at the segmental level, there appear to be differences in terms of the relative weightings placed on different cues. In the present study, junctural cues appeared to have the strongest effect on segmentation. In the $2 \times 2 \times 2$ analysis, above, partial η^2 values were reported as 0.985 for RTs, and 0.817 for accuracy – far higher than that of any other cue in the analysis. However, it is difficult to compare this effect with that of the PWC, since the values comes from a different type of analysis. Using η^2 values, rather than partial η^2, would result in a value of 0.58 for the RTs. In contrast, the comparison between *fapple* and *veef-apple* (the measure of PWC most akin to that used by Norris et al.) resulted in an η^2 value of 0.66 for RTs, whereas the comparison between *fapple* and *vee-fapple* resulted in an η^2 value of 0.21. Thus, depending on the particular measures used, the effects of junctural cues can be viewed as either stronger or weaker than that of the PWC, but clearly stronger than that of any other cue in the present study. That said, these weightings are likely based more on the clarity of acoustic information than on any strict ordering of cues. Thus, while junctural information was one of the most important cues in the present set of studies, its relative effectiveness may be tied to the individual speaker's patterns of pronunciation. If so, the extent to which listeners use one cue vs. the other will vary depending on the situation. This is, in essence, the argument put forth by Mattys et al. in their studies using background noise: when one cue becomes less useful in a particular context, other cues may play a larger role. We would suggest that such differences can be caused not only by noise, but also by the particular ways in which a given speaker produces different segments.

These findings suggest that, rather than having a tripartite hierarchy of segmentation cues, as suggested by Mattys et al., the process of segmentation involves a distinction between information present in the signal, and information based on prior lexical knowledge. Within the former, cues are each weighted according to their relative strength, but cannot be simply grouped into one or two stages of processing. Within the latter, information that represents an absolute such as the PWC is clearly stronger than probabilistic constraints.

Acknowledgments

The authors wish to thank Anne Schutte for recording the stimuli, and Rachael Dolezal, Isma Hussain, Ben Schnoor, Erica Stewart, Sarah Stilwill, Susan Timm, Phillippe Taborga, Andrea Tuttle and Brooke Werner for assistance testing participants, and Anne Cutler, John Kingston, and James McQueen for their comments on earlier drafts. Preparation of this manuscript was supported by a developmental assignment from the University of Iowa to the first author, by research grants HD37822-01 from NICHD and BCS99-07849 from NSF to the University of Iowa, and by NIDCD research grant R01-DC00219 to the University at Buffalo.

A. Supplementary material

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.jml.2010.11.004.

References

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements:

Evidence for continuous mapping models. *Journal of Memory & Language, 38*, 419–439.
Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13*, 221–268.
Christie, WM. Jr., (1974). Some cues for syllable juncture perception in English. *Journal of the Acoustical Society of America, 55*, 819–821.
Christie, W. M. Jr., (1977). Some multiple cues for juncture in English. *General Linguistics, 17*, 212–222.
Church, K. W. (1987). Phonological parsing and lexical retrieval. In U. H. Frauenfelder & L. K. Tyler (Eds.), *Spoken word recognition. Cognition special issues* (pp. 53–69). Cambridge, MA: MIT Press.
Cluff, M. S., & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 551–563.
Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.
Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133–163). Hillsdale, NJ: Erlbaum.
Content, A., Dumay, N., & Frauenfelder, U. H. (2000). The role of syllable structure in lexical segmentation in French: Helping listeners avoid mondegreens. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Spoken word access processes*. Nijmegen, The Netherlands: Max-Planck Institute for Psycholinguistics.
Cutler, A., Demuth, K., & McQueen, J. M. (2002). Universality versus language-specificity in listening to running speech. *Psychological Science, 13*, 258–262.
Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113–121.
Davis, M. H., Marslen-Wilson, W. D., & Gaskell, M. G. (2002). Leading up the lexical garden path: Segmentation and ambiguity in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 28*, 218–244.
Derwing, B. L. (1992). A "pause-break" task for eliciting syllable boundary judgments from literate and illiterate speakers: Preliminary results for five diverse languages. *Language and Speech, 35*, 219–235.
Dilley, L., Shattuck-Hufnagel, S., & Ostendorf, M. (1996). Glottalization of word-initial vowels as a function of prosodic structure. *Journal of Phonetics, 24*, 423–444.
Dukes, K. D., & Nakatani, L. H. (1976). *A study of acoustic cues related to juncture perception (Technical Memorandum (TM-76-1228-3))*. Murray Hill, NJ: Bell Laboratories.
Dumay, N., Frauenfelder, U. H., & Content, A. (2001). The role of the syllable in lexical segmentation in French: Word-spotting data. *Brain & Language, 81*, 144–161.
Dutton, D. (1992). The role of allophonic variation in the perception of word boundaries. Dissertation abstracts international, vol. 53, p. 4398.
Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America, 101*, 3728–3740.
Gaygen, D., & Luce, P. A. (2002). Troughs and bursts: Probabilistic phonotactics and lexical activation in the segmentation of spoken words in fluent speech. In McLennan, C. T., Luce, P. A., Mauner, G., & Charles-Luce, J. (Eds.), *University at Buffalo working papers on language and perception*, vol. 1, pp. 496–549.
Han, M. S. (1994). Acoustic manifestations of mora timing in Japanese. *Journal of the Acoustical Society of America, 96*, 73–82.
Hanulikova, A., McQueen, J. M., & Mitterer, H. (2010). Possible words and fixed stress in the segmentation of Slovak speech. *Quarterly Journal of Experimental Psychology, 63*, 555–579.
Hoard, J. E. (1966). Juncture and syllable structure in English. *Phonetica, 15*, 96–109.
Johnson, E. K., & Jusczyk, P. W. (2001). Word segmentation by 8-month-olds: When speech cues count more than statistics. *Journal of Memory and Language, 44*, 548–567.
Jusczyk, P. W., & Aslin, R. N. (1995). Infants' detection of the sound patterns of words in fluent speech. *Cognitive Psychology, 28*, 1–23.
Kirk, C. J. (2001). Phonological constraints on the segmentation of continuous speech. Doctoral dissertation, University of Massachusetts, Amherst. (2002, dissertation abstracts international – A, 62, 3366.)
Klatt, D. H. (1976). Linguistic uses of segmental duration in English: Acoustic and perceptual evidence. *Journal of the Acoustical Society of America, 59*, 1206–1221.
Klatt, D. H. (1980). Speech perception: A model of acoustic–phonetic analysis and lexical access. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 243–288). Hillsdale, NJ: Lawrence Erlbaum.
Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica, 5*(Suppl.), 5–54.
Luce, P. A., & Large, N. R. (2001). Phonotactics, density, and entropy in spoken word recognition. *Language and Cognitive Processes, 16*, 565–581.
Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing, 19*, 1–36.
Magnuson, J. S., Dixon, J. A., Tanenhaus, M. K., & Aslin, R. N. (2007). The dynamics of lexical competition during spoken word recognition. *Cognitive Science, 31*, 133–156.
Martin, J. G., & Bunnell, H. T. (1982). Perception of anticipatory coarticulation effects in vowel-stop consonant-vowel sequences. *Journal of Experimental Psychology: Human Perception and Performance, 8*, 473–488.
Mattys, S. L. (2004). Stress versus coarticulation: Toward an integrated approach to explicit speech segmentation. *Journal of Experimental Psychology: Human Perception and Performance, 30*, 397–408.
Mattys, S., & Jusczyk, P. W. (2001). Phonotactic cues for segmentation of fluent speech by infants. *Cognition, 78*, 91–121.
Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477–500.
McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology, 18*, 1–86.
McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21–46.
McQueen, J. M. (2005). Speech perception. In K. Lamberts & R. Goldstone (Eds.), *The handbook of cognition* (pp. 255–275). London: Sage.
McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of vocabulary. *Language and Cognitive Processes, 10*, 309–331.
McQueen, J. M., Otake, T., & Cutler, A. (2001). Rhythmic cues and Possible-Word Constraints in Japanese speech segmentation. *Journal of Memory and Language, 45*, 103–132.
Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America, 62*, 714–719.
Nakatani, L. H., & O'Connor-Dukes, K. (1979). *Phonetic parsing cues for word perception (Technical Memorandum TM-79-1228-4)*. Murray Hill, NJ: ATT Bell Laboratories.
Newman, R. S., Sawusch, J. R., & Luce, P. A. (1997). Lexical neighborhood effects in phonetic processing. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 873–889.
Newman, R. S., Sawusch, J. R., & Luce, P. A. (2005). Do postonset segments define a lexical neighborhood? *Memory & Cognition, 33*, 941–960.
Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition, 52*, 189–234.
Norris, D., McQueen, J. M., & Cutler, A. (1995). Competition and segmentation in spoken-word recognition. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 21*, 1209–1228.
Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The Possible-Word Constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191–243.
Norris, D., McQueen, J. M., Cutler, A., Butterfield, S., & Kearns, R. (2001). Language-universal constraints on speech segmentation. *Language and Cognitive Processes, 16*, 637–660.
Öhman, S. E. G. (1966). Coarticulation in VCV utterances: Spectrographic measurements. *Journal of the Acoustical Society of America, 39*, 151–168.
Port, R. F., Dalby, J., & O'Dell, M. (1987). Evidence for mora timing in Japanese. *Journal of the Acoustical Society of America, 81*, 1574–1585.
Quené, H. (1991). Word segmentation in meaningful and nonsense speech. *Paper presented at the 12th international congress of phonetic sciences, Aix-en-Provence, France, August 19–24*.
Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics, 20*, 331–350.
Ratcliff, R. (1993). Methods for dealing with reaction time outliers. *Psychological Bulletin, 114*, 510–532.
Reddy, R. (1976). Speech recognition by machine: A review. *Proceedings of the IEEE, 64*, 501–531.
Repp, B., Liberman, A., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance, 4*, 621–637.
Shatzman, K. B., & McQueen, J. M. (2006a). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68*, 1–16.

Shatzman, K. B., & McQueen, J. M. (2006b). The modulation of lexical competition by segment duration. *Psychonomic Bulletin & Review, 13*, 966–971.

Treiman, R., & Danis, C. (1988). Syllabification of intervocalic consonants. *Journal of Memory and Language, 27*, 87–104.

Umeda, N., & Coker, C. H. (1975). Subphonemic details in American English. In G. Fant & M. A. A. Tatham (Eds.), *Auditory analysis and perception of speech* (pp. 539–564). Hillsdale, NJ: Academic Press.

Umeda, N., & Coker, C. H. (1974). Allophonic variation in American English. *Journal of Phonetics, 2*, 1–5.

van der Lugt, A. H. (2001). The use of sequential probabilities in the segmentation of speech. *Perception & Psychophysics, 63*, 811–823.

Vitevitch, M. S. (2002). Influence of onset density on spoken-word recognition. *Journal of Experimental psychology: Human Perception and Performance, 28*, 270–278.

Vroomen, J., & de Gelder, B. (1995). Metrical segmentation and lexical inhibition in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 98–108.

White, L., Melhorn, J. F., & Mattys, S. (2010). Segmentation by lexical subtraction in Hungarian speakers of second-language English. *Quarterly Journal of Experimental Psychology, 63*, 544–554.