

# Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English

Taehong Cho<sup>a,b,\*</sup>, James M. McQueen<sup>b</sup>, Ethan A. Cox<sup>c</sup>

<sup>a</sup>Division of English Language and Literature, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul, 133-791, South Korea

<sup>b</sup>Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

<sup>c</sup>Department of Linguistics, Northwestern University, Illinois, USA

Received 4 August 2004; received in revised form 7 February 2006; accepted 27 March 2006

---

## Abstract

We explore the role of the acoustic consequences of domain-initial strengthening in spoken-word recognition. In two cross-modal identity-priming experiments, listeners heard sentences and made lexical decisions to visual targets, presented at the onset of the second word in two-word sequences containing lexical ambiguities (e.g., *bus tickets*, with the competitor *bust*). These sequences contained Intonational Phrase (IP) or Prosodic Word (Wd) boundaries, and the second word's initial Consonant and Vowel (CV, e.g., [tʃ]) was spliced from another token of the sequence in IP- or Wd-initial position. Acoustic analyses showed that IP-initial consonants were articulated more strongly than Wd-initial consonants. In Experiment 1, related targets were post-boundary words (e.g., *tickets*). No strengthening effect was observed (i.e., identity priming effects did not vary across splicing conditions). In Experiment 2, related targets were pre-boundary words (e.g., *bus*). There was a strengthening effect (stronger priming when the post-boundary CVs were spliced from IP-initial than from Wd-initial position), but only in Wd-boundary contexts. These were the conditions where phonetic detail associated with domain-initial strengthening could assist listeners most in lexical disambiguation. We discuss how speakers may strengthen domain-initial segments during production and how listeners may use the resulting acoustic correlates of prosodic strengthening during word recognition.

© 2006 Elsevier Ltd. All rights reserved.

---

## 1. Introduction

A key assumption of prosodic theories is that spoken utterances are produced in a hierarchically organized prosodic structure in which higher constituents are decomposed exhaustively into lower ones (M. Beckman, 1996; Beckman & Pierrehumbert, 1986; Jun, 1998; Nespor & Vogel, 1986; Pierrehumbert & Beckman, 1988; Selkirk, 1984, 1986; see Shattuck-Hufnagel & Turk, 1996, for a review). Prosodic structure is signaled in many ways. Fine-grained yet systematic phonetic variation in the production of individual speech sounds thus appears to be due in part to prosodic structure, or to prosodic context (Byrd, Kaun, Narayanan, & Saltzman, 2000; Cho, 2002, 2004, in press; Cho & Keating, 2001; De Jong, 1995; Edwards, Beckman, & Fletcher, 1991;

---

\*Corresponding author. Division of English Language and Literature, Hanyang University, 17 Haengdang-dong, Seongdong-gu, Seoul, 133-791, Korea. Tel.: +82 2 2220 0746; fax: +82 2 2220 0741.

E-mail address: [tcho@hanyang.ac.kr](mailto:tcho@hanyang.ac.kr) (T. Cho).

Fougeron, 2001; Fougeron & Keating, 1997; Jun, 1993; Keating, Cho, Fougeron, & Hsu, 2003). Furthermore, prosodic structure has been argued to constrain the phonological shapes of segments, either by forming different domains for application of phonological rules (Jun, 1998; Selkirk, 1986), or by licensing and neutralizing phonological contrasts depending either on the strengths of different prosodic positions or on the richness of the cues associated with those positions (Barnes, 2002; J. Beckman, 1998; Steriade, 1999). It has also been shown that segments at the edges of larger prosodic domains are produced so as to enhance phonological features in some phonetic dimensions (Cho, 2005; Cho & Jun, 2000). All of these studies support the view that prosodic structure is a principal organizational structure in speech production, and the view that prosody is a grammatical entity in its own right (Beckman, 1996). But it is not yet clear what roles prosodic structure plays in speech comprehension. In the present study, we therefore investigated the effect of prosodically conditioned phonetic variation on spoken-word recognition, in particular, the effect on lexical segmentation of prosodically driven phonetic detail in English domain-initial consonants.

### 1.1. Domain-initial strengthening

In their exploration of the phonetic correlates of the prosodic structure of utterances, researchers have traditionally paid particular attention to the ends (the right edges) of prosodic domains. For example, prosodic units larger than a (prosodic) word are generally defined in terms of the distribution of phrase-final intonational contrasts (e.g., Bolinger, 1970; Pierrehumbert, 1980; Pierrehumbert & Beckman, 1988) and final lengthening (e.g., Beckman, Edwards, & Fletcher, 1992; Cooper & Paccia-Cooper, 1980; Edwards et al., 1991; Klatt, 1975; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). There are, however, also durational and tonal correlates of prosodic structure in domain-initial position. Several recent studies have shown that domain-initial segments undergo so-called **prosodic strengthening**. This can be defined as a spatio-temporal expansion of articulation and a coarticulatory resistance. Both of these phenomena are proportional to the strengths of prosodic boundaries or to the degree of accentuation. For example, it has been demonstrated that the lip opening gesture in *ba#ba* (where # = a prosodic boundary) is generally larger and longer both domain-initially and domain-finally than domain-medially (Cho, 2002, *in press*); that the vocalic articulation at both edges in a sequence such as *ba#bi* shows greater resistance in vowel-to-vowel coarticulation spanning a larger than a smaller prosodic boundary (Cho, 2002, 2004); and that vowels /a,i/ are produced more peripherally before a higher prosodic boundary in a way to enhance frontedness for /i/ and backness/lowness for /a/ (Cho, 2005). Such evidence suggests that hierarchically organized prosodic structure is signaled at least in part by prosodic strengthening that is graded as a function of the level of the prosodic boundary.

Prosodic strengthening specifically associated with left edges of prosodic domains has become known as **domain-initial strengthening**. Domain-initial consonants are more strongly articulated than domain-medial consonants, and this effect is often cumulative, in that articulation is stronger (and longer) in domain-initial positions at each level in the prosodic hierarchy (Word, Intermediate Phrase, Intonational Phase) than in domain-medial positions at that level (Byrd et al., 2000; Cho, 2002, *in press*; Cho & Keating, 2001; Choi, 2003; Cole, Choi, Kim, & Hasegawa-Johnson, 2003; Fougeron, 2001; Fougeron & Keating, 1997; Jun, 1993; Keating et al., 2003; Pierrehumbert & Talkin, 1992; Tabain, 2003a, b). For example, voiceless aspirated stops, as in English and Korean, are generally produced with longer Voice Onset Times (VOTs) in domain-initial than in domain-medial positions (Cho & Jun, 2000; Choi, 2003; Cole et al., 2003; Cooper, 1991; Jun, 1993; Pierrehumbert & Talkin, 1992). Furthermore, coronal consonants are produced with larger linguopalatal contact in higher domains than in lower ones, as measured by electropalatography (Cho & Keating, 2001; Fougeron, 2001; Fougeron & Keating, 1997; Keating et al., 2003). Such evidence suggests that the prosodic hierarchy has direct phonetic correlates: the higher the prosodic constituent, the stronger the articulation of the initial segment.

### 1.2. Spoken-word recognition

The idea that domain-initial position has a special status in speech recognition is not new: It has often been considered to be an informationally rich locus in speech processing (see, e.g., Gow, Melvold, & Manuel, 1996; Nakatani & Dukes, 1977; Nooteboom, 1981; Quené, 1992; Shattuck-Hufnagel, 1986). In the Cohort model

(Marslen-Wilson, 1987; Marslen-Wilson & Welsh, 1978), for instance, word onsets play a critical role in determining which words are considered: Words which begin in the same way as the input word, and only those words, are considered as alternative lexical hypotheses. The precise function of the phonetic consequences of domain-initial strengthening on word onsets in speech perception, however, has not previously been examined. Researchers who have engaged in phonetic studies of domain-initial strengthening have merely speculated on its communicative functions. The simple assumption has been that the speaker signals prosodic structure via articulatory domain-initial strengthening, and the listener makes use of the acoustic consequences of this behavior in comprehension (see, e.g., Fougeron, 1999, for a review; Chapter 7 in Cho, 2002; cf. Harrington, 2003). More specifically, Fougeron and Keating (1997) discussed the benefits that listeners might receive from domain-initial strengthening, including assistance with lexical segmentation. They speculated that since domain-initial strengthening entails increased articulatory contrast between segments straddling a prosodic boundary, this contrast could contribute to marking that boundary, and thus help listeners to parse the incoming speech signal into words and phrases. Likewise, Cho and Jun (2000) interpreted the pattern of consonantal domain-initial strengthening that they observed as being related to the enhancement of phonological features and phonological contrasts, and hypothesized that these kinds of enhancements could ultimately facilitate word recognition by enhancing lexical distinctions.

The goal of the present study was therefore to investigate the role of domain-initial strengthening in comprehension directly. In so doing, we also provided additional acoustic data on consonantal strengthening in English from the single-speaker utterances used to make the experimental materials. More generally, we attempted to link phonetically oriented research on domain-initial strengthening with research in experimental psycholinguistics. One way of bridging this gap is to link (phonetic) models of articulatory control with (psycholinguistic) models of lexical access in speech production. In the general discussion we present an analysis of how the phonetic data on domain-initial strengthening informs theories of the relationship between phonological encoding and articulatory planning during speech production. Our primary theoretical focus, however, is on speech comprehension. While there are numerous aspects of speech comprehension that need to be explored in connection with domain-initial strengthening, we concentrated here on the role that phonetic strengthening plays in lexical segmentation. The reason for this choice was based on the fact that continuous speech lacks fully reliable cues to word boundaries (like the white spaces between written English words; see Norris, McQueen, Cutler, & Butterfield, 1997, for review). A central problem that listeners must thus solve during speech comprehension is that of finding discrete words in the quasi-continuous speech stream. The hypothesis that we tested here was that listeners use the acoustic correlates of domain-initial strengthening to help solve this segmentation problem. These correlates could be used to signal the location of possible word boundaries in continuous speech. Using the cross-modal identity priming task, we therefore investigated whether the acoustic-phonetic fine detail associated with a subset of domain-initial consonants in English (the voiceless fricative [s] and the voiceless stops [p,t,k]) is used by listeners to segment continuous speech into words. That is, we examined not simply whether listeners are sensitive to domain-initial strengthening in general, but, more specifically, how they might use such information in decoding continuous speech.

All current models of spoken-word recognition assume that as a listener hears an utterance, the words that are consistent with the input are activated, and that these multiple candidate words compete with each other for selection (see McQueen, 2005, for review). For instance, word candidates that begin in the same way (e.g., tickle and ticket) are activated as the initial part ([tʃ]) of the intended word (e.g., ticket) is being heard, and as more information becomes available over time in support of the intended word, the activation of the competitor word (e.g., tickle) is reduced until it eventually is no longer involved in the competition process. Inter-word competition through mutual inhibitory connections between word representations is a crucial component in word recognition models such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994). (See, e.g., Gaskell & Marslen-Wilson, 1997, and Luce & Pisoni, 1998, for different word recognition models in which there is still a lexical competition process, but one that is not implemented via inter-word inhibitory connections.) Lexical competition, however it may be implemented, helps to solve the segmentation problem given continuous speech input: As candidate words win the competition, word boundaries are “found” between them.

This competition process, however, is also modulated by cues which signal likely word boundaries. Metrical and phonotactic information appears to influence lexical segmentation in this way (e.g., McQueen, 1998;

Norris, McQueen, & Cutler, 1995). Furthermore, recent experimental results suggest that systematic, yet fine-grained phonetic detail, such as the non-contrastive duration of speech sounds, modulates lexical segmentation, as shown across a range of languages and experimental tasks (Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Davis, Marslen-Wilson, & Gaskell, 2002; Gow, 2001, 2002; Gow & Gordon, 1995; Quené, 1992; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Collina, Mazzetti, & Zoppello, 2000). Because acoustic cues to word boundaries are not fully reliable (Lehiste, 1972; Nakatani & Dukes, 1977), however, it is necessary to postulate a default segmentation mechanism that can operate successfully when such cues are not available (Norris et al., 1997). Lexical competition can serve this function. We therefore assume that fine-grained phonetic cues to word boundaries, when they are available in the speech input, act to modulate the competition process. We will examine here how the acoustic correlates of domain-initial strengthening interact with lexical competition during spoken-word recognition.

### 1.3. Prosody in lexical access

Several studies on the role of fine-grained acoustic–phonetic cues in comprehension have examined the effects of prosodic structures of different sizes on lexical access. Tabossi et al. (2000), for example, showed that the acoustic correlates of syllabic structure influence word recognition in Italian. In a fragment priming study, responses to a word which matched the syllable structure of the prime (e.g., *sil.vestre*, *sylvan*, after the prime [sil]) were facilitated, relative to control conditions, more so than responses to a syllabically mismatching target (e.g., *si.lenzio*, *silence*). This effect reversed when the prime was [si.l]. The effect was attributed to the influence of the acoustic signature of syllabic structure (in this case, the duration of the [l]) on lexical activation.

Other studies have examined the effect of cues to prosodic word (Wd) boundaries on lexical access (e.g., Gow & Gordon, 1995). One particular line of enquiry has examined the case of shorter words embedded in the onset of other, longer words (e.g., *cap* in *captain*). Using the cross-modal identity-priming paradigm, Davis et al. (2002) found evidence of more activation for the shorter words (e.g., *cap*) when listeners heard sequences where those words were intended by the speaker (e.g., *cap tucked*) than when they heard sequences containing the corresponding longer words (e.g., *captain*). The critical sequences (e.g., [kæp]) were acoustically longer when the speaker intended the monosyllabic word. These results suggest that differential acoustic cues in embedded syllables assist listeners in distinguishing monosyllabic words from the onsets of longer words. Salverda et al. (2003) argued that these acoustic cues were the consequence of a difference in prosodic structure (a Wd boundary after the monosyllabic word, but not after the first syllable of the polysyllabic word). Using an eye-tracking paradigm with cross-spliced Dutch materials, they demonstrated that activation of short words (e.g., *ham*, *id.*) in longer words (e.g., *hamster*, *id.*) was greater when the spliced first syllable had originated from the word *ham* than when it had originated from *hamster*, and showed that this effect appeared to be due to the duration of the spliced syllables. Salverda et al. therefore proposed that “a prosodic structure is built in parallel to the lexical analysis of the utterance and that the presence of segmental lengthening favors lexical candidates whose word boundaries are aligned with the predicted prosodic boundary” (p. 78).

Christophe et al. (2004) have recently investigated the role in spoken-word recognition of a prosodic boundary larger than that of the Wd. In many models of prosodic organization (e.g., Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; Selkirk, 1984), there are at least two prosodic domains larger than a word: the Phonological (or Intermediate) Phrase and then, at a yet higher level, the Intonational Phrase (IP). Christophe et al. examined Phonological Phrase boundaries. In word monitoring and phoneme detection tasks in French, they found evidence that lexical access for monosyllabic words (e.g., *chat*, *cat*) was faster when these words appeared in two-word sequences (e.g., *chat grincheux*, lit. *cat grumpy*) which spanned Phonological Phrase boundaries than when the sequences occurred within Phonological Phrases. Note that these sequences were temporarily lexically ambiguous: They were locally consistent with longer words which spanned the word (and Phrase) boundaries (e.g., *chagrin*, *sorrow*). Christophe et al. argued that listeners terminate pending lexical searches when they encounter Phonological Phrase boundaries, and therefore that lexical access in continuous speech is constrained such that it operates within the domain of the Phonological Phrase.

The acoustic correlates of different levels of prosodic structure also appear to help infants segment speech. Johnson (2003a, b) showed that 12-month old American infants preferred to listen to words (e.g., *lore*) that were matched to the word boundaries in two-word sequences (e.g., *toga # lore*) that they had been exposed to in a prior familiarization phase. No preference was found when the stimuli did not correspond to the word boundary in the two-word sequences that the infants had heard during familiarization (e.g., *toe # galore*). These results were taken to indicate that 12-month-olds use acoustic cues from prosodic structure (i.e., Wd boundaries) to segment speech into words, presumably as a precursor to word learning. Similarly, Gout, Christophe, and Morgan (2004) demonstrated that 13-month-old American infants preferred to turn their heads upon hearing a familiarized trochaic bisyllabic word (e.g., *paper*) when the word occurred within a sentence, but that this effect disappeared when the word straddled a Phonological Phrase boundary (e.g., *pay # permit*).

All these studies make clear that fine-grained acoustic-phonetic details stemming from a variety of prosodic structures place constraints on lexical access. To date, however, the influence on spoken-word recognition of phonetic details arising from IP boundaries has not been investigated. We examined that influence here. Another open issue about the role of higher-level prosodic boundaries in lexical access concerns the phonetic source of such effects. As mentioned above, prosodic structure is manifested in various phonetic phenomena. It is therefore critical to establish the extent to which these acoustic correlates are used by listeners during word recognition. Here, we focus on domain-initial strengthening, and include a detailed analysis of the acoustic-phonetic correlates of consonantal strengthening in our materials. The scope of the present study is thus confined to the effects on lexical segmentation of detailed phonetic outcomes that ensue from domain-initial strengthening associated with IP boundaries.

#### 1.4. Domain-initial strengthening in lexical segmentation

The primary hypothesis was that listeners would make use of the fine-grained acoustic-phonetic consequences of domain-initial strengthening in lexical segmentation. We tested this hypothesis using the cross-modal identity-priming paradigm. American English listeners heard American English sentences as they saw letter strings on a computer screen. Their task was to listen to the sentences and to decide whether or not the letter strings were real words. In the experimental trials, two-word sequences in the spoken sentences (e.g., *bus # tickets*, where # indicates a prosodic boundary), were paired with related visual target words, presented at the onset of the second word in the critical two-word sequences. In Experiment 1, the related targets were the second words in these sequences (e.g., *tickets*); in Experiment 2 they were the first words (e.g., *bus*). This manipulation allowed us to examine the effect of domain-initial strengthening on the recognition of the domain-initial word itself (Experiment 1), and to test whether it also influences the segmentation process, thereby affecting the recognition of other words (in this case the pre-boundary, domain-final words; Experiment 2). The influences were measured by comparing priming effects across conditions, that is, the relative speed and accuracy of responses to a visual target either when that word appeared in the auditory prime sentences (identical trials) or did not (unrelated trials).

The effect of domain-initial strengthening was tested by cross-splicing. In one version of each sentence (the same-spliced version), the initial CV sequence of the second word in the sequence (e.g., [tɪ] in *bus tickets*) was spliced from another version of that sentence. In a second version (the cross-spliced version), the initial CV of the second word was spliced from a matched sentence with a mismatched prosodic boundary. The initial CV was chosen for splicing, rather than just the onset consonant, given evidence (Cho, 2004, 2005, *in press*) that the effect of strengthening is not limited to the initial segment, but also spreads into the following segment. Furthermore, some acoustic correlates of consonants such as formant transitions often lie in the following vowel. Within each experiment, splicing was manipulated in two sub-experiments which differed only in whether the two-word sequence straddled an IP boundary or a Wd boundary. In Experiments 1a and 2a, the carrier sentences contained an IP boundary between the two critical words (e.g., *When you get on the bus, # tickets should be shown to the driver*). The cross-spliced versions contained CV word onsets that had originated from matched Wd-initial positions (e.g., from *John forgot to buy bus # tickets for his family*). The critical words in both types of sentence were not accented. (see Section 2.1.3 for further discussion on controlling accent). In Experiments 1b and 2b, the carrier sentences each contained a Wd boundary (they were in fact

tokens of the same sentences as were used for cross-splicing in the other set of sub-experiments), and the cross-spliced versions contained CV word onsets from segmentally matched IP-initial positions (i.e., from the carrier sentences used in Experiments 1a and 2a).

This set of conditions (summarized in Table 1) allowed us to test the effect of the acoustic correlates of a larger domain boundary (i.e., an IP boundary) in a context where a smaller Wd boundary would be expected—and the reverse—on the recognition of both the post- and pre-boundary words. An essential component of the design was that the critical two-word sequences were temporarily lexically ambiguous. The first word plus the beginning of the second word was always consistent with at least one other word (e.g., **bust** in **bus tickets**). The selection of items with these lexical competitors was intended to make it harder for listeners to segment the two-word sequences correctly and hence to increase the likelihood of obtaining evidence of a strengthening effect. Notice, however, that competition is likely to be fiercer between the pre-boundary word and the competitor, because the pre-boundary word was always embedded within the competitor (e.g., **bus** in **bust**), than between the competitor and the post-boundary word, which overlapped less (usually only on the last/first phoneme, e.g., the [t] in **bust** and **tickets**). It was possible that the strength of the lexical competition could modulate the effect of the prosodic manipulation.

It is also important to note that the degree of domain-initial strengthening may vary as a function of segment type. Fougeron (2001), for example, showed that the domain-initial strengthening effect on French consonants, as reflected in linguopalatal contact, is less robust for the fricative /s/ than for the stop /t/, just as the fricative /s/ generally exerts greater coarticulatory resistance to the influence of its context than stops do (e.g., Recasens, 1999). As Fougeron noted, this is presumably due to greater articulatory precision or “fewer articulatory and acoustic degrees of freedom” for /s/ than for /t/ (e.g., Shadle & Scully, 1995). Kim (2003a) also observed a lesser degree of articulatory domain-initial strengthening associated with /s/ in Korean in comparison with the patterns of domain-initial strengthening for Korean stops reported in Cho and Keating (2001). The acoustic consequences of articulatory domain-initial strengthening appear to differ between stops and nasals as well. Fougeron and Keating (1997) demonstrated that while linguopalatal contact and closure (seal) duration (as measured by EPG) for English /n/ is larger utterance-initially (and IP-initially preceded by a pause), the acoustic duration of /n/ is shorter utterance-initially than utterance-medially. Cho and Keating (2001) observed a similar pattern in Korean, and further showed that nasal energy generally decreases more domain-initially than domain-medially even when no pause is involved. This is in line with the diminution of nasal flow for IP-initial /n/ relative to that for initial /n/ in lower domains (e.g., Wd-initial but IP-medial /n/) which had been found in French (Fougeron, 2001). Overall, these results manifest more consistent and robust

Table 1  
Design and example stimuli

Experiment	Spoken prime sentence			Visual target word	
	Carrier sentence	Post-boundary CV Same-spliced	Post-boundary CV Cross-spliced	Identical	Unrelated
1a: IP-boundary sentences	“When you get on the bus, # tickets should be shown to the driver”	[tɪ] from IP-boundary	[tɪ] from Wd-boundary	tickets	company
1b: Wd-boundary sentences	“John forgot to buy bus # tickets for his family”	[tɪ] from Wd-boundary	[tɪ] from IP-boundary	tickets	company
2a: IP-boundary sentences	“When you get on the bus, # tickets should be shown to the driver”	[tɪ] from IP-boundary	[tɪ] from Wd-boundary	bus	mill
2b: Wd-boundary sentences	“John forgot to buy bus # tickets for his family”	[tɪ] from Wd-boundary	[tɪ] from IP-boundary	bus	mill

domain-initial strengthening patterns for stops than for non-stops. This led us to explore whether the hypothesized role of domain-initial strengthening in lexical segmentation varies depending on the type of consonant involved. Specifically, we compared sequences where the second word began with one of the three voiceless stops /p,t,k/ (e.g., *bus tickets*) with sequences where the second word began with the voiceless fricative /s/ (e.g., *ray signal*, with the competitor *race*).

## 2. Experiment 1

We first examined whether the acoustic manifestation of domain-initial strengthening influences recognition of the domain-initial word itself. In Experiment 1a, we examined recognition of words in post-IP-boundary position (e.g., *tickets* in *When you get on the bus, # tickets should be shown to the driver*). In the same-spliced condition, the initial CVs of those words contained acoustic cues associated with IP-initial position. In the cross-spliced condition, these CVs were substituted with the acoustic cues associated with Wd-initial position. We predicted that there would be an identity-priming effect: faster and/or more accurate responses to the visual targets (e.g., *tickets*) as a sentence containing that word was heard (specifically, at the onset of the auditory word *tickets*) than as an unrelated sentence was heard (e.g., [Davis et al., 2002](#); [Gow, 2002](#); [Spinelli et al., 2003](#)).

Experiment 1b employed a complementary materials set to examine the effects of the same splicing manipulation as in 1a, this time on words in post-Wd-boundary position. The recognition of these words (e.g., *tickets* in *John forgot to buy bus # tickets for his family*) was examined in a same-spliced condition in which the initial CV of the post-boundary word contained acoustic information that was associated with the beginning of a word following a Wd boundary, and a cross-spliced condition in which the same CV portion had been replaced with a CV excised from IP-initial position. Here again we predicted priming effects in both splicing conditions.

The critical question was whether the size of the priming effects would vary as a function of the splicing manipulations. If acoustic cues associated with an IP boundary assist in the recognition of the domain-initial word, the priming effect in Experiment 1a should be stronger in the same-spliced condition than in the cross-spliced condition. In Experiment 1b, in contrast, priming should be stronger in the cross-spliced than in the same-spliced condition. These two tests also allowed us to examine whether any effect of domain-initial strengthening would be contextually conditioned. In the same-spliced condition of Experiment 1a the cues associated with IP-initial position appeared in a context where those cues were appropriate (the unspliced remainder of each sentence was also consistent with an IP boundary). In the cross-spliced condition of Experiment 1b, however, the cues associated with IP-initial position appeared in a context where those cues were inappropriate (the unspliced remainder of each sentence was consistent with a Wd boundary).

Finally, it was also possible to ask whether, irrespective of the splicing manipulation, words in the critical two-word sequences would be recognized more easily when straddling an IP boundary (Experiment 1a) than when straddling a Wd boundary (Experiment 1b). If the pre-boundary cues associated with a larger prosodic boundary assist in segmentation, then the local ambiguity in the two-word sequence (e.g., *bus* vs. *bust* in *bus # tickets*) should be resolved more rapidly across an IP boundary than across a Wd boundary, resulting in improved recognition of the post-boundary words in Experiment 1a relative to in Experiment 1b. That is, the priming effect should be stronger in the former experiment.

### 2.1. Method, experiment 1a

#### 2.1.1. Participants

Forty-eight student volunteers from the University of Arizona were paid for their participation. They were all speakers of American English, with no known hearing problems.

#### 2.1.2. Materials

Forty-eight pairs of English sentences like those in [Table 1](#) were constructed. The same critical two-word sequence (e.g., *bus tickets*) appeared in both sentences within a pair. In one sentence there was an IP boundary between the two critical words. In the other sentence there was a Wd boundary at that location. In

Experiment 1a, the IP-boundary sentences were the experimental carriers. (In Experiment 1b, the Wd-boundary sentences were the experimental carriers.)

In each two-word sequence the first word plus the beginning of the second word was another word or the beginning of another word (e.g., **bust** in **bus tickets** or **partner** in **part names**). The initial consonant of the second word was either a voiceless stop ([p], [t], or [k]), the fricative [s] or the nasal [n]. Eighteen sentence pairs had a voiceless stop, another 18 pairs had [s], and another 12 pairs had [n]. Only 12 sentence pairs for [n] could be constructed given the joint constraints on stimulus selection. Note that the [n] sentence pairs were excluded from the analyses. The visual target words associated with the 48 experimental sentences were the second words in the critical two-word sequences (either, on identical trials, exactly the same word as appeared in the sentence, e.g., **tickets** for the **bus tickets** sentence, or, on unrelated trials, a word from a different experimental sentence, e.g., **company** for the **bus tickets** sentence). Lists of prime sentences, target words and representative competitors are given in Appendix A.

A further 112 filler and 10 practice sentences were constructed. Each contained a two-word sequence with an IP boundary. For 48 of them, matched sentences containing the same two-word sequence, but separated by a Wd boundary, were constructed. Half of the 48 filler IP-boundary sentences with matched Wd-boundary sentences were used to make same-spliced fillers, and half to make cross-spliced fillers. The remaining 64 filler sentences and the practice sentences were unspliced. For half of the 48 spliced filler sentences (12 cross-spliced and 12 same-spliced), non-word targets were made that were phonologically related to the second word in the critical pair (e.g., for the sequence **bill dates**, the non-word **dakes**, which has the same onset, vowel and final consonant as **dates**). Here, as in all other experiments we report, non-words were orthotactically and phonotactically legal in English. For the other 24 spliced filler sentences and a further 32 unspliced fillers, non-word targets that were phonologically unrelated to the spoken sequences were made (e.g., for the sentence containing the sequence **small chairs**, the non-word **fep**). Real words that were phonologically and semantically unrelated to the spoken sequences were selected as targets for the other 32 unspliced fillers (e.g., for the sequence **poultry fried**, the word **treatment**). These additional materials were thus constructed such that splicing (should it be detected by listeners) and formal similarity between prime and target words could not be used by listeners as signals for a strategic “YES” response. Five practice sentences were paired with real-word targets (two identical to the second word in the sequence, e.g., **forks** for the sequence **attack forks**, and three unrelated); the other practice sentences were paired with non-word targets (two phonologically related, e.g., **clating** for the sequence **party clearing**, and three unrelated).

### 2.1.3. Procedure

Multiple tokens of each sentence were recorded in a sound-damped booth by a male native speaker of American English who was naive about the experiment’s purpose. Two versions (same-spliced vs. cross-spliced) of each experimental sentence were then made using the Praat speech editor (<http://www.praat.org>). The same carrier token of each IP sentence was used in each version. The same-spliced version was made by splicing the initial CV of the domain-initial (post-boundary) word (e.g., [t] of **tickets**) from another token of that IP sentence into the carrier sentence. The cross-spliced version was made by splicing into the same carrier the initial CV of the domain-initial word from a token of the segmentally matched Wd boundary sentence. The onset of each CV was defined as (a) the release of the closure (for stops), (b) the beginning of the high-frequency frication noise (for [s]) or (c) the beginning of the nasal murmur (for [n]). The offset of each CV was defined, in all cases, as the offset of the vocalic energy associated with the first vowel in the domain-initial word. Comparable splice points were selected across each pair of CVs. All splices were made at zero-crossings in the waveform. In some cases where the CV portion was followed by a liquid coda (/r,l/), the liquid portion was included, partly due to difficulties of segmenting the vowel and the coda liquid and partly to prevent the resulting stimuli from sounding unnatural. (We will nonetheless refer to the spliced portions as CVs.) The same splicing procedure was applied to the filler sentences with matched Wd sentences (24 were same-spliced and 24 were cross-spliced).

Note that for the stops, since the onset of the CV was defined as the release of the closure, the duration of the silent stop closure was not included in the splicing. This means that stop closure lengthening was not manipulated across conditions. There were two inter-related reasons for this. The first was practical: For voiceless stops, it is not absolutely clear, especially in IP-initial position, where in the signal the stop closure



begins. Splicing the entire silent period before the stop release is likely to include a silent period which can not be attributable unambiguously to consonantal articulation, and hence not to domain-initial consonantal strengthening (i.e., a pause associated with the IP boundary that is independent of stop production). The second reason was theoretical: Although lengthening of domain-initial stops has been considered as one of the most robust acoustic consequences of domain-initial strengthening, and domain-initial strengthening effects are generally local to the beginning of the consonant and tend to wane moving into the vowel, some results on suggest that there are other phonetic cues that are present after the consonantal constriction (e.g., VOT, stop burst energy, spectral characteristics of the release, vowel formants, and vowel-to-vowel coarticulatory resistance, see Cho, 2004, 2005; Cho & Keating, 2001; Choi, 2003; Cole et al., 2003; Tabain, 2003a). We therefore decided to eliminate differences in stop closure duration between the splicing conditions. This allowed us to test, for stop-initial CVs, the specific effects of the post-release acoustic–phonetic events associated with domain-initial strengthening.

To ensure that any observed patterns were derived from domain-initial strengthening, other prosodic factors had to be controlled. First, in order to avoid possible confounding effects from accentuation (or nuclear pitch accent), the speaker was instructed to accentuate words other than the critical two words (e.g., **bus tickets**) in all the experimental, filler and practice sentences. The controlling of accent was particularly important because the lexical categories of the critical two words were different from token to token (e.g., noun–noun (e.g., **bus tickets**), adjective–noun (e.g., **high tide**), verb–noun (e.g., **buy salmon**). These syntactic category differences have consequences for which of the two words is the syntactic head, and hence could cause different default accent patterns across stimuli. These in turn might have influenced processing of the two-word sequences. The words which speakers were asked to accentuate were highlighted in upper case in the recording script (e.g., **When YOU get on the bus, tickets should be shown to the DRIVER and John FORGOT to buy bus tickets for his FAMILY**). The result was that the critical words were de-accented. Second, in order to elicit uncontroversial IP boundaries, all sentences were constructed such that the expected IP boundaries were always aligned with a major syntactic boundary (e.g., a clause boundary); this boundary was always marked with a comma in the recording script. Out of 4–6 repetitions of each sentence, the two best tokens (one for the carrier sentence and one for splicing) were selected. (Note also that before the actual recording, the speaker practiced the speech material for 30 min, focusing on the boundary type and accent placement.) On the basis of the guidelines of the English Tones and Break Index system (ToBI, Beckman & Elam, 1997; Silverman et al., 1992), the selection criteria were that the expected prosodic boundary did indeed fall between the two words in the critical sequence (e.g., **bus # tickets**) and that these two words were not accented. Stimulus selection was made with the aid of pitch and intensity displacement measures generated by Praat. Third, the two versions of each experimental sentence were equated for F0 differences on the spliced CV portions of the post-boundary word using the PSOLA resynthesis method, based on mean F0 values of the paired spliced CVs. This F0 neutralization was done in order to eliminate any potential influences of F0 differences between IP- and Wd-initial CVs on spoken-word recognition (cf. Kim, 2003b). All other phonetic cues for a relevant prosodic boundary remained intact, such that lengthening and boundary tones, as well as consonant duration for IP-boundary sentences, were available to the listeners. (Note that all the IP-boundary sentences had the same tonal pattern: The speaker consistently produced L-H% sequences before the boundary.) Informal inquiry of the participants confirmed that they did not notice any unnaturalness in the speech materials.

Each listener heard all the sentences once, and saw all of the targets once. Splicing (same- vs. cross-spliced sentences) and relatedness (visual target identical to the domain-initial word in the critical sequence or unrelated to it) were counterbalanced across four lists which were presented to four groups of listeners. Experimental sentences and targets were paired (e.g., the auditory sentences with **bus tickets** and **mill company**, and the visual targets **tickets** and **company**) such that each list contained 12 trials in each of the four counterbalanced conditions without within-list repetition of either the sentences or the targets, but also so that across the experiment as a whole, all stimuli (sentences and targets) appeared in all four conditions.

Each list also contained all the filler trials, mixed in random order with the experimental trials. Any given participant therefore saw 48 experimental and 32 filler word targets and 80 non-word targets. Twenty four of the words (i.e., 15% overall) were identical to the second (domain-initial) word in the critical two-word

sequence being heard; 24 non-words were phonologically related to the second word in these sequences. Furthermore, each type of spliced sentence (same- and cross-spliced) was just as likely to be paired with a word as with a non-word target.

Participants were tested individually in a quiet room. Sentences were presented over headphones at a comfortable listening level. Targets appeared for 1 s in lower case on the screen of a laptop computer. Each target appeared on the screen in synchrony with the onset of the domain-initial segment (i.e., at the onset of the spliced CV portion of the post-boundary word). Participants were asked to listen to the sentences, and to decide as quickly and as accurately as possible whether the target letter strings were real English words. They had to respond by pressing one of two buttons, labeled “YES” and “NO.” Positive responses were made with their preferred hand. The computer clock was triggered in synchrony with the onset of the target presentation on the screen and stopped either when the response button was pressed or at the acoustic offset of the sentence if no response was made. Response latencies and errors were collected.

All participants received the practice trials, followed by one of the four lists. They were also informed that they would be given a comprehension test at the end of the experiment. This was to ensure that the participants would listen carefully to the auditory materials, and to draw their attention to the meaning of the sentence, and hence to discourage them from attending only to the visual targets. This pen-and-paper test comprised 16 printed sentences; half of them had been presented auditorily, and half were similar to the presented materials but new. Immediately after the main session, each participant had to judge whether the printed sentences had appeared in the main part of the experiment. They made, on average, 10.0 (63%) correct responses.

## 2.2. Method, experiment 1b

### 2.2.1. Participants

A further 52 student volunteers from the University of Arizona were paid for their participation. They were all speakers of American English, with no known hearing problems.

### 2.2.2. Materials

The materials were based closely on those used in Experiment 1a. The 48 experimental sentences were those with Wd boundaries that were used for cross-splicing in Experiment 1a. There were again 48 filler sentences that were spliced, 64 that were unspliced, and 10 practice sentences. The spliced filler sentences were those that were matched to the spliced filler sentences in Experiment 1a (i.e., they had the same critical two-word sequences, but had Wd boundaries rather than IP boundaries). For the experimental and spliced filler sentences, the visual targets (words and non-words, respectively) were the same as in the matched sets of stimuli in Experiment 1a. The remaining sentence materials (unspliced fillers and practice sentences) were new. All of these had a Wd boundary near the middle of the sentence. As before, half of the unspliced fillers were paired with phonologically and semantically unrelated target words (e.g., for the sequence **special glue**, the target **hamsters**) and the other half were paired with phonologically unrelated non-words (e.g., for the sequence **horrible monsters**, the target **spreap**). The practice sentences were paired to related and unrelated word and non-word targets in the same way as in Experiment 1a.

### 2.2.3. Procedure

The additional sentences were recorded by the same speaker, during the Experiment 1a recording session. The same editing procedure was used for splicing the stimuli, but now such that all carrier sentences had Wd boundaries, and were either made by splicing the initial CV of the post-boundary word from a different token of the same sentence or from a matched IP-boundary sentence. All 48 experimental sentences were both same- and cross-spliced, while again half of the spliced filler sentences were same-spliced and half were cross-spliced. The same procedure as in Experiment 1a was adopted for counterbalancing materials across four lists, and for running the experiment. In the comprehension test, listeners made, on average, 11.1 (69%) correct responses.

## 2.3. Results

### 2.3.1. Acoustic analyses

The CV portions of the experimental stimuli that were used for splicing were analyzed to test for durational and amplitude differences between those CVs originating from IP-boundary contexts and those originating from Wd-boundary contexts. The acoustic measurements included: duration of the entire CV; consonant duration for /s/ and /n/; VOT for stops (recall that stop closures were not spliced); vowel duration; and peak amplitude (in dB) during the stop burst, during frication for /s/, during nasalization for /n/, and during the vowel. The results are given in Appendix B.

There was strong evidence of the acoustic consequences of domain-initial strengthening. The duration of the entire spliced CV was significantly longer when the CVs originated from IP- than from Wd-boundary contexts, both for the CVs used in the IP boundary carrier sentences (Experiment 1a) and for the CVs used in the Wd-boundary carriers (Experiment 1b). These durational differences were carried by the initial C's; those for the V's alone were not significant. Another important finding was that peak amplitude during the vowel was significantly greater for CVs in IP-initial position than for CVs in Wd-initial (IP-medial) position, again for the materials in both experiments. This confirms that the domain-initial strengthening effect is not limited to the initial consonant, but also spreads into the following vowel (see Cho, 2004, 2005, *in press*).

Some differences among consonant types were also observed. The fricative /s/ showed significant domain-initial consonantal lengthening in Experiments 1a and 1b. But VOT for stops, the temporal measure relevant to consonant production, revealed no significant strengthening effect (though the mean differences were in the expected direction). With respect to the consonantal peak amplitude measure, however, only the stops showed a domain-initial strengthening effect. Peak amplitude during the stop burst was significantly larger for CVs from IP boundaries than for CVs from Wd boundaries, while no such effect was found for peak amplitude during /s/. Both stop and fricative CVs showed significant peak amplitude effects on the following vowel. The nasal /n/ showed the least evidence of acoustic–phonetic cues to domain-initial strengthening. A significant effect on nasal duration (in both experiments) was observed, but there were no strengthening effects on peak amplitude during the /n/ or during the vowel following the /n/.

In short, there were substantial acoustic phonetic differences between IP-initial CVs and IP-medial CVs. The spliced CVs therefore carried durational and amplitude differences due to domain-initial strengthening. However, the effect was not consistent across consonant types. For stops, despite the exclusion of closure durations from the splicing manipulation, there were still robust cues of domain-initial strengthening as reflected in greater peak amplitude during the stop burst and during the following vowel for IP-initial CVs than for Wd initial (i.e., IP medial) CVs. The fricative /s/ showed significantly longer fricative duration and greater peak-vocalic amplitude in the IP-initial CVs. The nasal /n/ showed only a consonantal duration effect.

### 2.3.2. Experiment 1a results

Lexical decision reaction times (RTs) in this and all subsequent experiments were measured from the onset of the visual presentation of the target words. For each subject, RTs above or below 2.5 standard deviations from the grand mean for that subject were treated as errors (2.47%). Responses in the conditions in which the targets began with an /n/ were excluded from the main analyses in this and all subsequent experiments, for three reasons. First, there were only 12 such items. Second, their exclusion allowed for a balanced design, with equal proportions of items in which the target began with either a stop or a fricative (18 of each). Given the differences in the acoustic measures between stops and fricatives, it was important to test for different perceptual effects across these two specific consonant types. Third, and most importantly, the acoustic analyses revealed that there were only relatively weak domain-initial strengthening effects in the /n/ materials. A summary of mean RTs and error rates in the /n/ conditions is given in Appendix C.

Mean RTs and error rates for the stop- and fricative-initial targets are shown in Fig. 1. The results were evaluated with three-way repeated measure analyses of variance (ANOVAs) with the factors Consonant Type (targets beginning with stops vs. fricatives), Splicing (initial CV of the post-boundary word in the prime sentence same-spliced vs. cross-spliced) and Relatedness (post-boundary prime in sentence and target identical vs. unrelated). F-values were evaluated for analyses with subjects (F1) and with items (F2) as the repeated

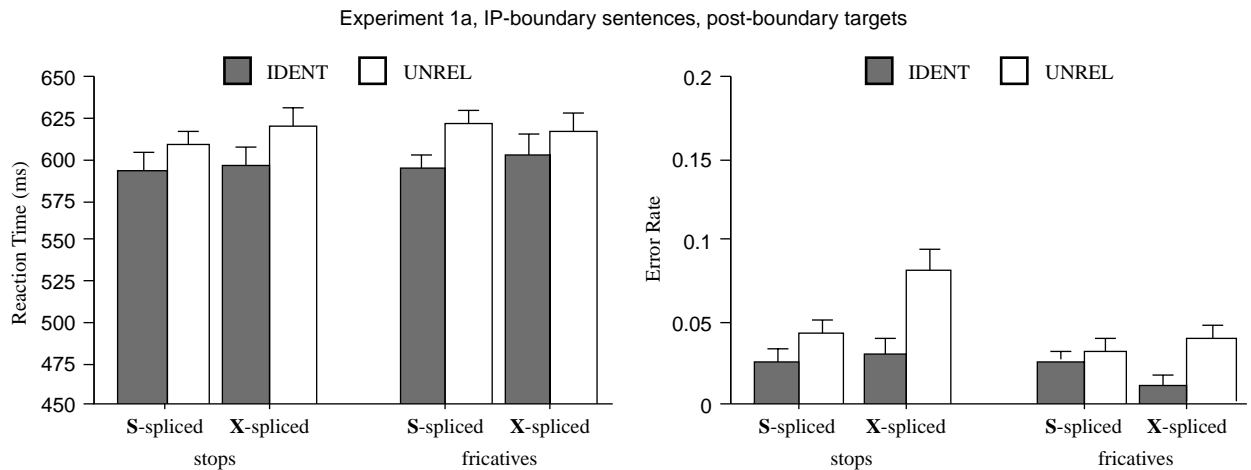


Fig. 1. Mean RTs (in milliseconds) and error rates (proportion of incorrect responses) for visual lexical decisions in Experiment 1a. Targets corresponded to the post-boundary (domain-initial) word in two-word sequences (e.g., bus # tickets) in the prime sentences. There was an IP boundary between the two words in these sequences (e.g., When you get on the bus, # tickets should be shown to the driver). Data are plotted for responses to targets which were identical (IDENT) and unrelated (UNREL) to material in the prime sentences and which began either with a stop or a fricative. The initial CV of the post-boundary word in each sentence was either same-spliced (S-spliced) from another token of the same sentence or cross-spliced (X-spliced) from a matched sentence with the same two-word sequence, but one spanning a Wd boundary.

measure. In the by-subjects ANOVA, all three factors were within-subjects; in the by-items ANOVA, Consonant Type was a between-items factor and Splicing and Relatedness were within-items factors.

As illustrated in Fig. 1, the results revealed an identity-priming effect on RTs: Listeners hearing IP-boundary priming sentences were faster to respond to identical targets (e.g., to tickets as “bus # tickets” was heard, where # = IP) than to unrelated targets (e.g., to tickets as “mill # company” was heard): 597 vs. 619 ms, on average;  $F(1, 44) = 13.34$ ,  $p < 0.001$ ,  $F(2, 34) = 5.28$ ,  $p < 0.05$ ). There was neither a main effect of Consonant Type ( $F(1, 44) < 1$ ;  $F(2, 34) < 1$ ) nor a significant interaction between Relatedness and Consonant Type ( $F(1, 44) < 1$ ;  $F(2, 34) < 1$ ). Similarly, there was neither a main effect of Splicing ( $F(1, 44) < 1$ ;  $F(2, 34) < 1$ ) nor a significant interaction between Splicing and Relatedness ( $F(1, 44) < 1$ ;  $F(2, 34) < 1$ ). Finally, there was no significant three-way interaction (Splicing  $\times$  Relatedness  $\times$  Consonant Type,  $F(1, 44) < 1$ ;  $F(2, 34) < 1$ ). In other words, there was a priming effect on RTs, but this effect was not modulated by the acoustic detail in the initial CV of the post-boundary word in the prime sentence. Priming effects were observed whether this CV originated from an IP-initial context or a Wd-initial context, and whether the CV included a stop or a fricative.

With respect to error rates (see Fig. 1), there was a main effect of Relatedness in the analysis by subjects ( $F(1, 44) = 6.96$ ,  $p < 0.05$ ), but only a trend in the analysis by items ( $F(2, 34) = 3.29$ ,  $p < 0.1$ ). As with the RTs, there were no other significant main effects or interactions. The overall mean error rate was 3.5%. The pattern in the errors was therefore in line with the RT pattern: Listeners’ responses tended to be more accurate when they responded to a related target than to an unrelated target, but no other manipulated factors modulated this effect.

### 2.3.3. Experiment 1b results

RTs above or below 2.5 standard deviations from each subject’s grand mean were again treated as errors (2.88%). Mean RTs and error rates are shown in Fig. 2. The results were analyzed in the same way as in Experiment 1a. The findings were similar. Listeners were faster to respond to identical targets (e.g., to tickets as “bus # tickets” was heard, but where # is now a Wd boundary) than to unrelated targets (565 vs. 584 ms, on average;  $F(1, 46) = 16.95$ ,  $p < 0.001$ ;  $F(2, 34) = 17.45$ ,  $p < 0.001$ ). Unlike in Experiment 1a, there was a main effect of Consonant Type; this effect, however, was significant only in the analysis by subjects ( $F(1, 46) = 7.45$ ,  $p < 0.01$ ;  $F(2, 34) = 1.43$ ,  $p > 0.2$ ). There was therefore a trend for responses to be faster

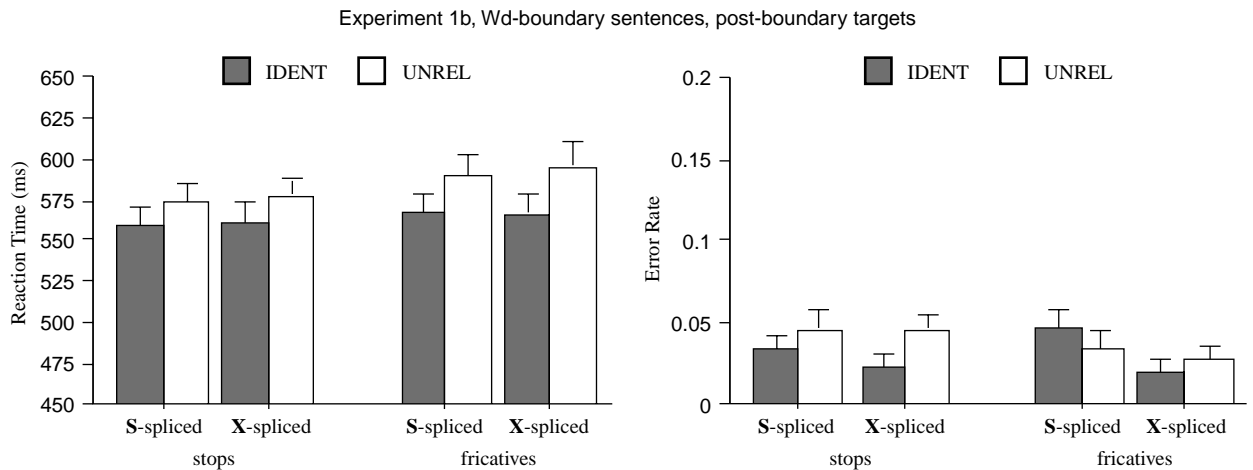


Fig. 2. Mean RTs (in milliseconds) and error rates (proportion of incorrect responses) for visual lexical decisions in Experiment 1b. Targets corresponded to the post-boundary (domain-initial) word in two-word sequences (e.g., *bus # tickets*) in the prime sentences. There was a Wd boundary between the two words in these sequences (e.g., *John forgot to buy bus # tickets for his family*). Data are plotted for responses to targets which were identical (IDENT) and unrelated (UNREL) to material in the prime sentences and which began either with a stop or a fricative. The initial CV of the post-boundary word in each sentence was either same-spliced (S-spliced) from another token of the same sentence or cross-spliced (X-spliced) from a matched sentence with the same two-word sequence, but one spanning an IP boundary.

when targets were stop-initial than when they were fricative-initial. However, this factor showed no significant interaction with Relatedness, as in Experiment 1a ( $F(1, 46) = 1.14$ ,  $p > 0.2$ ;  $F(1, 34) < 1$ ). As in Experiment 1a, neither a main effect of Splicing nor a significant interaction involving the Splicing factor was found (Splicing:  $F(1, 46) < 1$ ,  $F(1, 34) < 1$ ; Splicing  $\times$  Relatedness:  $F(1, 46) < 1$ ,  $F(1, 34) < 1$ ; Splicing  $\times$  Relatedness  $\times$  Consonant Type:  $F(1, 46) < 1$ ;  $F(1, 34) < 1$ ). There were no significant effects in the error rates. The overall mean error rate was 3.6%.

#### 2.3.4. Combined analyses

We also conducted combined analyses of the RT data to see whether the priming effect would be modulated by the type of prosodic boundary (IP vs. Wd) in the prime sentences. This factor was a between-subject factor in the by-subjects analysis but a within-item factor in the by-items analysis. The main effect of Boundary Type was marginal by subject ( $F(1, 90) = 3.56$ ,  $p = 0.062$ ) and significant by item ( $F(1, 34) = 67.82$ ,  $p < 0.0001$ ; mean RTs in IP vs. Wd sentences: 611 vs. 575 ms). Crucially, however, the Boundary Type factor did not interact reliably with Relatedness or any other factor. This suggests that, independent of domain-initial strengthening and relatedness, targets tended to be recognized more slowly when the sentences contained IP boundaries.

#### 2.4. Discussion

As we predicted and in line with other cross-modal identity-priming studies (e.g., Davis et al., 2002; Gow, 2002; Spinelli et al., 2003), we found a priming effect in both versions of Experiment 1. Listeners were faster to respond, for example, to *tickets* upon hearing “*bus # tickets*”) than to *tickets* upon hearing “*mill # company*”. Contrary to our predictions, however, we found no influence of our splicing manipulation on the size of the priming effect, neither in the IP-boundary sentences in Experiment 1a nor in the Wd-boundary sentences in Experiment 1b. The manipulation of the acoustic consequences of domain-initial strengthening in the post-boundary words in the prime sentences therefore did not influence lexical decision performance. This in turn suggests that the acoustic correlates of domain-initial strengthening did not significantly influence the segmentation and recognition of the post-boundary words, both when the critical two-word sequences straddled an IP boundary and when they straddled a Wd boundary. In both cases it would appear that there

was sufficient information for successful recognition of the domain-initial word (as measured by the systematic priming effect across all four splicing conditions), and hence there was no additional recognition benefit to be had when IP-initial cues were present (in the same-spliced conditions in Experiment 1a and the cross-spliced conditions in Experiment 1b). Nor was there any decrement in recognition performance when IP-initial cues were absent (in the cross-spliced conditions in Experiment 1a and the same-spliced conditions in Experiment 1b).

Although our acoustic analyses revealed differences between the domain-initial stops and the domain-initial fricatives, there were no robust effects of consonant type on word recognition performance. The only indication of such a difference was the trend towards faster responses to stop-initial targets than to fricative-initial targets in Experiment 1b, which was significant only by subjects. Crucially, there were no interactions of consonant type with relatedness, nor three-way interactions of consonant type, relatedness, and splicing. There was therefore no evidence that the acoustic correlates of domain-initial strengthening in stop consonant contexts had a different effect on word recognition from the correlates associated with the fricative /s/. Finally, we found no reliable evidence for the influence of boundary type on the recognition of the domain-initial words. The combined analyses showed that the priming effect was no larger with IP-boundary sentences than with Wd-boundary sentences. But we did observe that, overall, the visual targets, identical and unrelated alike, were recognized more slowly in Experiment 1a (with IP-boundary sentences), than in Experiment 1b (with Wd-boundary sentences). Any interpretation of this between-subjects comparison must be treated with caution. One possible interpretation is that, because the sentences in Experiment 1a were more complex, clause-level comprehension was more demanding than in Experiment 1b, and lexical decisions to the visual targets were thus correspondingly slower.

At first glance, these results appear to challenge the claim that word recognition in continuous speech is influenced by the acoustic correlates of prosodic structure. Given that the correlates of Wd boundaries (Salverda et al., 2003) and of Phonological Phrase boundaries (Christophe et al., 2004) both seem to modulate lexical access, one would surely predict that those associated with larger prosodic domains would also have an effect on word recognition. There is a good reason, however, why Experiment 1 may have failed to detect a domain-initial strengthening effect. As we have already argued, word recognition entails a process by which multiple candidate words are incrementally and simultaneously activated in proportion to their similarity to the speech input, and a process of competition among those candidates. In such a framework, the strengthening of the initial syllable of the domain-initial word (e.g., the [t] in *tickets*) may increase bottom-up support for that word, but also, to an equal extent, that for other (cohort-like) competitors that begin in the same way (e.g., *tickles*). Since the acoustic correlates of domain-initial strengthening therefore do not assist in lexical disambiguation about the post-boundary word (i.e., they can not help the listener select *tickets* and reject *tickles*), they may have no role to play in recognition of that word.

This does not mean, however, that domain-initial strengthening has no role to play in lexical segmentation. As a lexically ambiguous fragment such as [bʌst] in *bus # tickets* is being heard, the strengthened [t] is likely to be identified as a word onset rather than as a coda. That is, even though the strengthened [t] may not tell the listener what the domain-initial word is, it does indicate that there is a new prosodic domain, beginning before the [t]. It is therefore possible that the recognition of the pre-boundary word (e.g., *bus*) could be facilitated by the strengthened [t], since it provides evidence against competitors which straddle the prosodic boundary (e.g., *bust*). In Experiment 2, we tested whether domain-initial strengthening influences lexical disambiguation in this way.

### 3. Experiment 2

Experiment 2 was a mirror image of Experiment 1. Lexical decisions were made to pre-boundary words (e.g., *bus* in *bus # tickets*) instead of post-boundary words. The experiment was otherwise identical, with again two sub-experiments which differed only with respect to the type of boundary (IP vs. Wd) in the experimental sentences. The auditory prime sentences with IP boundaries that were used in Experiment 1a were reused in Experiment 2a, and those with Wd boundaries that were used in Experiment 1b were reused in Experiment 2b.

We tested whether lexical decisions to domain-final words (e.g., *bus*) are facilitated as listeners hear a sequence containing that word (the identical condition, e.g., *bus # tickets*), relative to the unrelated condition

(see Table 1). Of specific interest was whether this priming effect would be more robust when the initial CV of the following domain-initial word contained stronger domain-initial strengthening cues (i.e., when the CV came from an IP boundary rather than from a Wd boundary). Our predictions were as follows. If the presence of rich pre-boundary cues before an IP boundary (e.g., boundary tones and final lengthening) play a pivotal role in disambiguating temporary lexical ambiguities, strengthening cues in the post-boundary word should have little impact on resolving the ambiguity in IP-boundary sentences, leading to no modulation of the priming effect by the splicing manipulation in Experiment 2a. In contrast, Wd-boundary sentences (as in Experiment 2b) lack robust pre-boundary cues. Under these conditions, the strengthened initial CV of the post-boundary word could play a stronger role in lexical segmentation. We therefore predicted that the splicing manipulation would modulate the priming effect in Experiment 2b. This experiment thus tests the conditions where phonetic detail associated with domain-initial strengthening could assist listeners most in word recognition. Unlike in Experiment 1, there is a lexical ambiguity to be resolved involving the prime word (e.g., *bus* vs. *bust*), and, unlike in Experiment 2a, there are no robust pre-boundary cues available to resolve this ambiguity. The strengthening of the [tʃ] in *bus # tickets*, for example, could therefore assist in recognition of *bus*, even though, as Experiment 1b showed, this information did not help listeners recognize *tickets*.

### 3.1. Method

#### 3.1.1. Participants

One hundred new volunteers (48 in Experiment 2a and 52 in Experiment 2b) from Ohio State University and the University of Arizona were paid to take part. They were all speakers of American English, with no known hearing problems.

#### 3.1.2. Materials and procedure

The auditory materials and the procedures were the same as in Experiment 1. The only changes were in the visual targets. The first words in the critical two-word spoken sequence in each experimental sentence now served as targets (once again, in both the identical and unrelated conditions). For half of the 48 pairs of spliced filler sentences (i.e., the 24 spliced IP-boundary fillers used in Experiment 2a, and the matched set of spliced Wd-boundary fillers used in Experiment 2b) new non-word targets were made that were phonologically related to the first word in the critical pair (e.g., for the sequence *bill dates*, the non-word *bim*, which has the same onset and vowel as *bill*, rather than, as in the previous experiment, the non-word *dakes*). For the other spliced filler sentences, all the unspliced fillers and the practice sentences, the same unrelated targets (words or non-words) were used as in the corresponding previous sub-experiment (i.e., the same unrelated filler targets in Experiment 2a as in Experiment 1a, and the same in 2b as in 1b). The visual targets still appeared aligned in time with the onset of the post-boundary words—if the visual targets were aligned in time with the onset of the pre-boundary word, the effect of the phonetic detail might not be observable, give that the critical phonetic detail was manipulated at the onset of the post-boundary word. In the post-experiment comprehension test, participants in Experiments 2a and 2b made on average, respectively, 11.6 (73%) and 11.9 (74%) correct recognition responses.

### 3.2. Results

#### 3.2.1. Experiment 2a results

The prime sentences were IP-boundary sentences. RTs above or below 2.5 standard deviations from each subject's mean were again treated as errors (2.82%). Responses to one item with an error rate greater than 20% were excluded from the analysis. For the remaining items, the overall mean error rate was 5.8%. Mean RTs and error rates are shown in Fig. 3. The results were evaluated in the same way as in Experiment 1.

There was a priming effect on RTs (549 vs. 589 ms, on average;  $F(1, 44) = 73.41$ ,  $p < 0.001$ ,  $F(1, 33) = 25.53$ ,  $p < 0.001$ ) as well as on error rates (3% vs. 8%, on average;  $F(1, 44) = 22.69$ ,  $p < 0.001$ ;  $F(1, 33) = 10.34$ ,  $p < 0.005$ ). Listeners' responses were thus faster and more accurate when they responded to an identical visual target (to *bus* as they heard *bus # tickets*, # = IP boundary) than to an unrelated visual target (to *bus* as they heard *mill # company*). No other effects were fully significant. In RTs, there was a trend

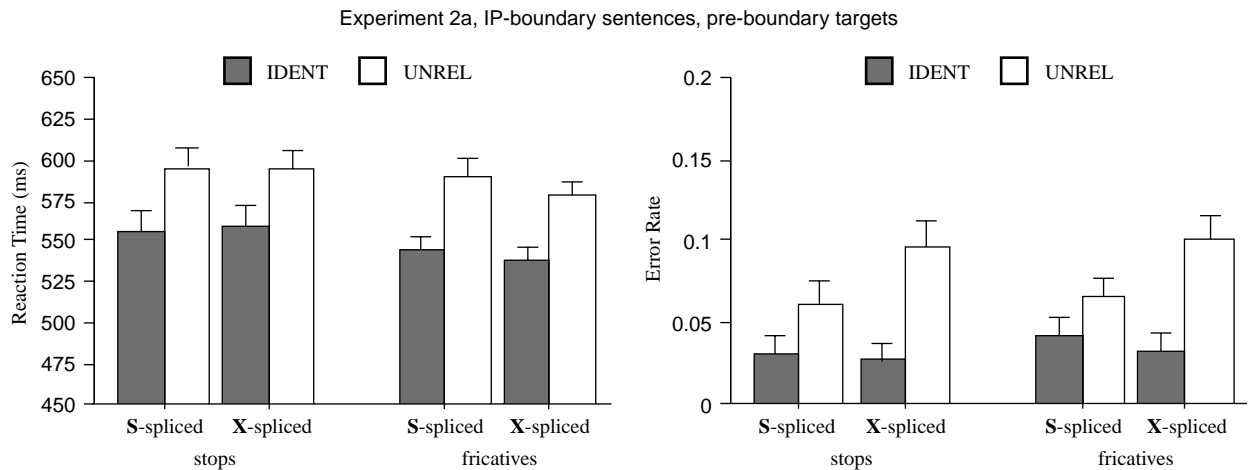


Fig. 3. Mean RTs (in milliseconds) and error rates (proportion of incorrect responses) for visual lexical decisions in Experiment 2a. Targets corresponded to the pre-boundary (domain-final) word in two-word sequences (e.g., *bus # tickets*) in the prime sentences. There was an IP-boundary between the two words in these sequences (e.g., *When you get on the bus, # tickets should be shown to the driver*). Data are plotted for responses to targets which were identical (IDENT) and unrelated (UNREL) to material in the prime sentences and which began either with a stop or a fricative. The initial CV of the post-boundary word in each sentence was either same-spliced (S-spliced) from another token of the same sentence or cross-spliced (X-spliced) from a matched sentence with the same two-word sequence, but one spanning a Wd boundary.

toward an effect of Consonant Type ( $F(1, 44) = 9.268, p < 0.005$ ;  $F(1, 33) = 1.431, p = 0.24$ ). Otherwise, all main effects and interactions, in both RTs and errors, were not significant (all  $F_s < 1$  for all cases except for the interaction between Consonant Type and Relatedness in RTs,  $F(1, 44) = 1.137, p > 0.2$ , and the interaction of Consonant Type and Splicing in RTs,  $F(1, 44) = 1.141, p > 0.2$ ). The priming effect therefore did not vary either as a function of the type of consonant in the domain-initial word (stop vs. fricative) or as a function of the splicing manipulation (i.e., target responses did not vary depending on whether the onset of the post-boundary word had been spliced from IP-initial position or from Wd-initial position).

### 3.2.2. Experiment 2b results

The prime sentences were Wd-boundary sentences. RTs above or below 2.5 standard deviations from each subject's mean were once again treated as errors (2.88%). Responses to two items with error rates greater than 20% were excluded from the analysis. For the remaining items, the overall error rate was 6.4%. Mean RTs and error rates are given in Fig. 4.

As in Experiment 2a, there was a priming effect in both RTs and error rates (RTs, 558 vs. 585 ms, on average;  $F(1, 45) = 27.35, p < 0.001$ ,  $F(1, 32) = 15.96, p < 0.001$ ; error rates, 4% vs. 9%, on average;  $F(1, 45) = 24.26, p < 0.001$ ;  $F(1, 32) = 5.48, p < 0.05$ ). Crucially, however, this priming effect was modulated by the splicing manipulation. Although there was no main effect of Splicing in either RTs or error rates (all  $F_s < 1$ ), there was a significant Relatedness by Splicing interaction in RTs ( $F(1, 45) = 7.71, p < 0.01$ ;  $F(1, 32) = 4.91, p < 0.05$ ), but not in errors (both  $F_s < 1$ ). The priming effect in RTs was larger in the cross-spliced condition (in which the initial CV of the post-boundary word had been spliced from IP-initial position) than in the same-spliced condition. Planned pairwise comparisons showed that the priming effect was significant both in the cross-spliced condition (effect size: 39 ms,  $t(48) = 5.96, p < 0.001$ ;  $t(33) = 4.12, p < 0.001$ ) and in the same-spliced condition (effect size: 16 ms,  $t(48) = 2.79, p < 0.01$ ;  $t(33) = 2.20, p < 0.05$ ). The differential priming effect was further confirmed by  $\eta^2$ -statistics (by-subjects,  $\eta^2 = 0.425$  vs. 0.139, by-items,  $\eta^2 = 0.340$  vs. 0.128, for the cross- and same-spliced conditions respectively).

Responses to targets after primes where the post-boundary word began with a fricative tended to be faster than after primes where the post-boundary word began with a stop, but this effect was only significant by subjects (563 vs. 580 ms, on average;  $F(1, 45) = 11.44, p < 0.005$ ,  $F(1, 32) = 1.40, p > 0.2$ ). More importantly, however, unlike in previous experiments, there was a significant Consonant Type  $\times$  Relatedness interaction in



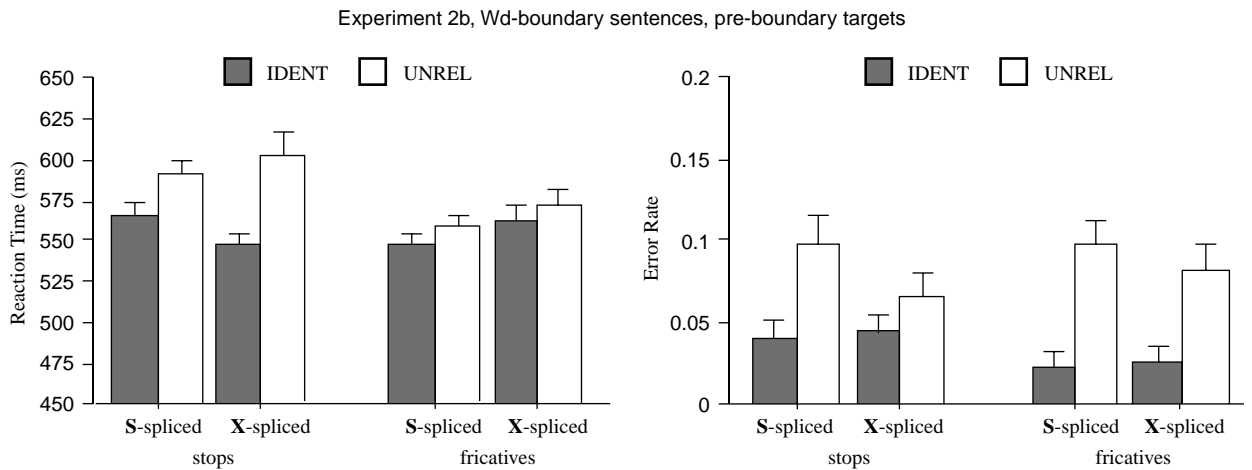


Fig. 4. Mean RTs (in milliseconds) and error rates (proportion of incorrect responses) for visual lexical decisions in Experiment 2b. Targets corresponded to the pre-boundary (domain-final) word in two-word sequences (e.g., *bus # tickets*) in the prime sentences. There was a Wd boundary between the two words in these sequences (e.g., *John forgot to buy bus # tickets for his family*). Data are plotted for responses to targets which were identical (IDENT) and unrelated (UNREL) to material in the prime sentences and which began either with a stop or a fricative. The initial CV of the post-boundary word in each sentence was either same-spliced (S-spliced) from another token of the same sentence or cross-spliced (X-spliced) from a matched sentence with the same two-word sequence, but one spanning an IP boundary.

response latencies in the analysis by subjects ( $F(1, 45) = 9.09$ ,  $p < 0.005$ ), and a trend in the analysis by items ( $F(2, 32) = 3.46$ ,  $p = 0.072$ ). As illustrated in Fig. 4, the priming effect was more robust when the post-boundary word began with a stop (560 vs. 601 ms, on average, for related vs. unrelated trials;  $t(48) = 4.77$ ,  $p < 0.001$ ,  $t(15) = 4.68$ ,  $p < 0.001$ ) than when the post-boundary word began with a fricative (557 vs. 568 ms, on average;  $t(48) = 4.68$ ,  $p < 0.001$ ;  $t(17) = 1.5$ ,  $p > 0.1$ ). There was, however, no significant three-way interaction in RTs between Splicing, Relatedness and Consonant Type ( $F(1, 45) = 1.48$ ,  $p > 0.2$ ,  $F(2, 32) < 1$ ). In other words, the differential priming effect as a function of Splicing was independent of Consonant Type. No other effects in the RT or error rate analyses were significant.

### 3.2.3. Combined analyses

As in Experiment 1, we also conducted combined analyses of the RTs and error rates from Experiments 2a and 2b, in order to test whether the priming effect would interact with boundary type. Unlike in Experiment 1, the RT analysis showed a significant interaction between Relatedness and Boundary Type, by subjects only ( $F(1, 89) = 3.87$ ,  $p < 0.05$ ;  $F(2, 32) < 1$ ). The priming effect was larger for IP-boundary sentences (effect size, 40 ms;  $t(48) = 9.10$ ,  $p < 0.001$ ;  $t(33) = 5.07$ ,  $p < 0.001$ ) than for Wd-boundary sentences (effect size, 27 ms;  $t(47) = 5.75$ ,  $p < 0.001$ ;  $t(33) = 3.86$ ,  $p < 0.001$ ). This asymmetry was confirmed by  $\eta^2$ -statistics (by subjects,  $\eta^2 = 0.625$  vs. 0.378, by items,  $\eta^2 = 0.436$  vs. 0.333, for IP-boundary vs. Wd-boundary sentences). The priming effect was therefore more robust in IP-boundary sentences (Experiment 2a) than in Wd-boundary sentences (Experiment 2b). The three-way interaction of Relatedness, Boundary Type and Splicing was not significant in the RT analysis ( $F(1, 89) = 3.90$ ,  $p = 0.051$ ;  $F(2, 32) = 1.77$ ,  $p > 0.1$ ). In the error rates, there was neither a significant two-way interaction between Relatedness and Boundary Type (both  $F$ s  $< 1$ ) nor a robust three-way (Relatedness by Boundary Type by Splicing) interaction ( $F(1, 89) = 5.01$ ,  $p < 0.05$ ;  $F(2, 32) = 2.47$ ,  $p > 0.1$ ). Finally, there was no main effect of Boundary Type either by subject or by item, in contrast to Experiment 1.

### 3.3. Discussion

In Experiment 2a, in which the critical two-word sequence straddled an IP boundary, we found that listeners were in general faster to respond to related than to unrelated targets. But there was neither a splicing effect nor an interaction between the splicing and relatedness factors. Simply put, the presence or absence of

the acoustic consequences of domain-initial strengthening in the post-boundary spoken words did not influence performance on the visual targets, which, in related trials, matched the pre-boundary spoken words. In accordance with our predictions, therefore, there was no evidence that the acoustic cues to domain-initial strengthening in the second word in a two-word sequence spanning an IP boundary influenced recognition of the pre-boundary word. It appears that domain-final (pre-boundary) cues in the word before an IP boundary (e.g., domain-final lengthening and/or a boundary tone on *bus* in *bus # tickets*) were sufficiently powerful to allow listeners to segment the sequence correctly (and thus rapidly rule out spurious candidate words such as *bust*). The presence of domain-initial strengthening cues in the initial CV of the following word thus had no observable effect on lexical segmentation.

In Experiment 2b, however, in which the critical two-word sequence straddled a Wd boundary, the priming effect interacted significantly with the splicing manipulation. In both the same- and cross-spliced conditions, listeners were faster and more accurate on identical than on unrelated targets. The latency effect, however, was stronger in the cross-spliced than in the same-spliced condition. We suggest that the acoustic-phonetic cues associated with domain-initial strengthening present in the cross-spliced sequences helped listeners to segment the critical sequence (i.e., helped them rule out competitors such as *bust* in *bus # tickets*). In contrast, when no strengthening cues were available (i.e., in the same-spliced sequences) but competitors were still present, listeners were slower to recognize the pre-boundary word, attenuating the priming effect.

The combined analyses showed no main effect of Boundary Type in contrast to Experiment 1. This contradicts the possible interpretation we offered for the finding in Experiment 1 that responses were slower in Experiment 1a (IP-boundary sentences) than in Experiment 1b (Wd-boundary sentences). If the complexity of clause-level comprehension were responsible for that earlier difference, we ought to have replicated the finding in Experiment 2. The effect we found in Experiment 1 may therefore be due instead to some other difference between subject groups. The combined analysis also showed that the priming effect in RTs when the target corresponded to a prime word before an IP boundary was more robust than when it corresponded to a prime before a Wd boundary. This suggests that, in addition to the effects of domain-initial strengthening cues, other information about the presence of a prosodic boundary can constrain lexical access (Christophe et al., 2004; Salverda et al., 2003). It appears that the pre-boundary acoustic cues to an IP boundary provide more support for a boundary after a word (e.g., *bus* in *bus # tickets*)—and hence act more strongly to penalize spurious candidate words (e.g., *bust*)—than the cues associated with an upcoming Wd boundary.

## 4. General discussion

### 4.1. Domain-initial strengthening in speech processing

We have shown that an American English speaker systematically distinguishes IP-initial consonants from Wd-initial consonants, and that listeners use the acoustic correlates of these differences as they recognize words in continuous American English utterances. First, in keeping with the cross-linguistic phonetic literature, our acoustic analyses of the speech materials produced by the naive speaker found evidence for domain-initial strengthening for both the voiceless stops and the fricative [s]. Both types of consonant showed greater peak amplitude during the vowel in IP-initial position than in Wd-initial position. These findings suggest that the initial strengthening effect is not limited to the consonant itself, but also spreads into the next vowel, in accordance with other recent findings (Cho, 2004, 2005, *in press*). Furthermore, only fricatives showed longer consonant duration (i.e., friction duration) in the IP-boundary context. That is, there was no significant difference in VOT between stops in the two prosodic contexts. The non-significance of this difference appears at first glance to be incompatible with previous results (e.g., Choi, 2003; Cole et al., 2003; Pierrehumbert & Talkin, 1992) which have shown that VOT in English is significantly longer phrase-initially than phrase-medially. Recall, however, that in the present study sentential accent (narrow focus) fell on words other than the target sequence in the sentence. The lack of a lengthening effect on VOT in unaccented initial positions, as in the present materials, can therefore be accounted for by the possibility that boundary effects are crucially enhanced by the presence of pitch-accent (e.g., White, 2002). Nevertheless, the VOT difference in the present materials was in the predicted direction, and stops (but not fricatives) were associated with greater peak amplitude during the consonant itself (i.e., during the stop burst). Though the data patterns were thus

somewhat different for stops and fricatives, both types of segment displayed clear evidence of several acoustic correlates of domain-initial strengthening.

Second, our series of cross-modal identity-priming studies has shown that listeners are sensitive to these acoustic correlates. More specifically, we have shown that listeners use this kind of prosodically-driven phonetic detail in resolving lexical ambiguities in continuous speech segmentation. Experiment 2b (with Wd-boundary carrier sentences) showed that as American English listeners heard a temporarily ambiguous sequence such as *bus tickets* (with the lexical competitor *bust*), the phonetic detail in the initial CV of the second word influenced recognition of the first word. There was stronger identity priming of target responses (e.g., to *bus*) when the initial CV of the second word originated from an IP-boundary context (e.g., the [tɪ] from *When you get on the bus, # tickets should be shown to the driver*) than when it originated from another token of the carrier sentence and hence from a Wd-boundary context (e.g., the [tɪ] from *John forgot to buy bus # tickets for his family*). We argued that the competition caused by the lexical ambiguity (e.g., between *bus* and *bust*) made it difficult for listeners to recognize the intended word, and that under these conditions the acoustic correlates of domain-initial strengthening could help listeners derive the correct segmentation (*bus* is aligned with the boundary cued by the prosodic detail, but *bust* is misaligned).

Our series of priming experiments also showed that listeners' apparent use of the acoustic consequences of domain-initial strengthening was limited to the situation where those acoustic cues could be of significant benefit to lexical segmentation. Thus, while the splicing manipulation modulated the priming effect in Experiment 2b (with Wd-boundary carriers and pre-boundary targets), there was no such interaction in Experiments 2a, 1b or 1a. We propose that there was no interaction in Experiment 2a because pre-boundary prosodic information was sufficient for listeners to resolve the lexical ambiguity. In Experiment 2a, the carrier sentences had IP boundaries (e.g., *When you get on the bus, # tickets should be shown to the driver*). Pre-boundary lengthening, boundary tones and greater coarticulatory resistance on the IP-final word may all have contributed to rejection of the competitor word (*bust*) and recognition of the intended word. No additional effect of the post-boundary cues could thus be detected. But there must be a different reason why there was no splicing by relatedness interaction in Experiment 1b, in which recognition of the post-boundary words (e.g., *tickets*) was measured in Wd-boundary carrier sentences. Strengthening of the initial syllable of these words should increase the bottom-up support not only for those words themselves but also, to an equal extent, for their cohort-like competitors (e.g., *tickles, ticking and ticker* for *tickets*). Strengthening therefore did not differentially benefit the recognition of the post-boundary words. Both of these factors (the presence of pre-boundary cues in IP-boundary carriers and the lack of a recognition benefit specific to the post-boundary word) may have prevented detection of a splicing by relatedness interaction in Experiment 1a.

One might argue that our results suggest that the detail associated with domain-initial strengthening has no role to play in normal word recognition. After all, the only condition where a benefit was measured was in the unnatural situation where an IP-initial CV was spliced into a Wd-boundary context (Experiment 2b). This conclusion is unwarranted, however. Prosodic strengthening can occur naturally in this kind of situation (i.e., in the vicinity of a Wd boundary)—strengthening due to accentuation or focus. (We thank an anonymous reviewer for this observation.) [Cooper and Paccia-Cooper \(1980: Ch. 6\)](#), for example, showed that palatalization across a Wd boundary within a phrase is less likely to occur when the post-boundary word receives narrow focus. Prosodic strengthening takes place in such contexts ([Cho, 2004, 2005, in press](#)). Thus, if the phonetic effects of domain-initial strengthening and accentuation are comparable, the present demonstration of a benefit to spoken-word recognition due to prosodically driven phonetic detail may reflect processing that would naturally take place when accentuation occurs in a Wd-boundary context. It would be wrong to conclude, however, that listeners only use strengthening cues in such situations. Lexical segmentation depends on a multitude of factors, including the effective set of lexical competitors (e.g., whether words have cohort competitors, and whether the current input is consistent with two close competitors such as *bus* and *bust*) and the availability of prosodically driven word boundary cues (e.g., pre- and post-boundary cues). Listeners are likely to use all the information that is available in a given situation to recognize words in continuous speech ([Norris et al., 1997](#)). Our results are consistent with this view. We were able to detect an independent contribution of domain-initial strengthening cues in a situation where other constraints on segmentation were controlled to be relatively weak (i.e., when there was strong competition between the intended words and their competitors, and where there were no strong pre-boundary cues). But

this does not mean that listeners do not use strengthening cues in other environments. Our results are consistent with the view that the phonetic detail associated with IP-initial consonants is used, along with other constraining information, in continuous speech segmentation.

#### 4.2. Prosodic boundary effects on spoken-word recognition

We also asked whether words in the critical two-word sequences were in general processed more efficiently when straddling IP boundaries (Experiments 1a and 2a) than when straddling Wd boundaries (Experiments 1b and 2b). It was hypothesized that the local ambiguity in the two-word sequence (e.g., *bus* vs. *bust* in *bus # tickets*) would be resolved more rapidly across an IP boundary than across a Wd boundary, resulting in improved recognition of the pre- and post-boundary words. In Experiment 2, in which the recognition of the pre-boundary word was measured, the priming effect was indeed stronger when the crucial two-word sequences straddled IP boundaries than when they straddled Wd boundaries. This was evident in the interaction between boundary and relatedness of the combined by-subjects analysis, though not in the combined by-items analysis. Planned pairwise comparisons and  $\eta$ -statistics confirmed that the priming effect was larger in Experiment 2a than in Experiment 2b. That is, recognition of the pre-boundary words appeared to be better at an IP boundary than at a Wd boundary. Further support for this claim comes from the failure to observe an initial strengthening effect in Experiment 2a. As we argued in the previous section, the pre-boundary cues in an IP-boundary context are strong enough to mask any effect of post-boundary strengthening cues on recognition of the pre-boundary word.

Taken together, these results are consistent with two recent proposals. The facilitatory effect could be attributable to the termination of lexical searches when a phrasal boundary, such as a Phonological Phrase or an IP boundary, is encountered (Christophe et al., 2004), or it could be driven by a preference for lexical candidates whose boundaries are aligned with a prosodic boundary (Salverda et al., 2003). These proposals were based on data in French and Dutch, respectively, so the present study has extended this line of investigation to English. It appears that pre-boundary acoustic cues to IP boundaries, relative to those to Wd boundaries, provide more support for the end of the pre-boundary word (e.g., *bus* in *bus # tickets*) and at the same time mismatch with the spurious candidate words (e.g., *bust*) more severely.

One might ask, however, why the boundary effect was not evident in the recognition of post-boundary words. In the combined analyses in Experiment 1, where the targets were the post-boundary words, we found no evidence for a boundary type by relatedness interaction. This null result may be due to the activation of multiple cohort-like candidate words in post-boundary position. Even if the segmentation of the two-word sequence were facilitated with the aid of strong acoustic–phonetic boundary cues, resulting in stronger activation of pre-boundary words, this need not also benefit the post-boundary words, since the evidence that there is a word onset at the boundary does not benefit recognition of a particular post-boundary word over that of its competitors. This evidence offers equal support for all cohort competitors of the post-boundary word. It appears, therefore, that prosodic cues to word boundaries primarily influence recognition of competing words when there is ambiguity about the location of the boundary (e.g., in the competition between *bus* and *bust*), and have less effect on the recognition of competing words which agree on the location of the boundary (e.g., *tickets* and *tickles*). In other words, prosodic information acts principally to resolve ambiguities about the placement of word boundaries.

#### 4.3. The effect of intrinsic phonetic properties

We found some evidence for a stop-fricative asymmetry in Experiment 2b. In Wd-boundary sentences, the priming effect on the pre-boundary words was larger when the post-boundary words began with stops than when they began with the fricative /s/. The interaction of consonant type and relatedness was significant by subjects, and marginal by items. Pairwise comparisons and  $\eta$ -statistics confirmed, however, that the priming effect observed for sentences containing voiceless stop onsets (/p,t,k/) was more robust than the effect observed for sentences with /s/-onsets.

This effect was independent of the domain-initial strengthening effect, indicated by the failure to observe a significant three-way interaction of relatedness, splicing and consonant type. To the extent this statistical null

effect holds, possible differences in the extent to which stops and fricatives may be strengthened domain-initially thus appear to be irrelevant. Although Fougeron (2001) measured such articulatory differences in French (comparing /t/ and /s/), and although our acoustic analyses indicated that the pattern of strengthening effects was different as a function of manner of articulation, we found no evidence that the domain-initial acoustic correlates were somehow stronger for the stops than for the fricatives. While it of course remains possible that quantitative differences between stops and fricatives in the extent of articulatory strengthening do exist in American English, we found no evidence that such differences influence word recognition.

We suggest that the global difference in the size of the priming effect between the stop- and fricative-initial conditions, though independent of the domain-initial strengthening effect, reflects differences in the informational value of the critical consonants. A major difference between the intended two-word sequences and the competing words in our materials is that all the intended-word sequences have the critical consonant in the onset of the second word whereas most of the competing words have it as their final coda consonant. This is true for the stops and /s/ (e.g., for the stops, bus[t] vs. bus [t]ickets, boa[t] vs. bow [t]ie; for the fricative, pea[ce] vs. pea [s]oup, ra[ce] vs. ray [s]ignal). The stops and the fricative /s/ have differential allophonic distributions between onset and coda positions, however.

Consider the stops first. There is a wide range of allophony for stops in word-final (or syllable-final) position. For example, the word-final stop in boat may be produced with or without a release (and/or aspiration) and with or without glottalization, resulting in one of the following five categories: [t<sup>h</sup>], [t], [t<sup>l</sup>], [ʔt] and [ʔ] (Ladefoged, 2000). But word-initial stops are always released into the following vowel, and this release is accompanied by aspiration, as opposed to word-final stops whose phonetic realization is allophonically more variable. Release and aspiration is thus more likely to be associated with initial than with final stops. The consistent release and aspiration in the stops in the initial CVs of the present speech materials was therefore a strong (but probabilistic) allophonic cue for lexical segmentation. Several other studies have shown that allophonically governed phonetic detail can influence segmentation performance (Dumay, Frauenfelder, & Content, 2000; Kirk, 2000; Smith, 2004; Smith & Hawkins, 2000; Spinelli et al., 2003; Tabossi et al., 2000). Stop release and aspiration in the initial CV may therefore have assisted listeners in Experiment 2b to disambiguate the stop-bearing two-word sequences. Furthermore, there is an onset-coda asymmetry for stops in the degree of articulatory strengthening. Keating, Wright, and Zhang (1999) demonstrated that word-initial English stops are produced with stronger articulation (as measured by EPG linguopalatal contact) than word-final stops are, just as domain-initial stops are more strongly articulated than domain-medial stops. It is therefore possible that there is acoustic-phonetic detail present in the speech signal that favors domain-initial stops, which listeners could use as additional cues to word onsets.

These differences may well be weaker for the fricative /s/. For word-initial onset vs. word-final coda /s/, in contrast to the stops, no impressionistic allophonic differences have been noted. Instrumental techniques have suggested, in addition, that differences between onset and coda /s/ are more limited than those found for stops. Indeed, in an EPG study, Byrd (1996a) found no significant spatial and temporal differences between word-initial and word-final /s/. Keating et al. (1999) also found no spatial expansion for word-initial /s/ relative to word-final /s/, though /s/ was longer word-initially than word-finally. As discussed earlier, the fricative /s/ is generally characterized as having “fewer articulatory and acoustic degrees of freedom” (Shadle & Scully, 1996), and shows less segmental and prosodic contextual variation (Fougeron, 2001; Kim, 2003a; Recasens, 1999). This is not to say that there are no prosodically driven effects on perception for fricatives, however. Shatzman and McQueen (2006), in a recent eye-tracking study in Dutch, have shown that durational differences between word-final /s/ and word-initial /s/ in ambiguous sequences such as eens pot (“once jar”) and een spot (“a spotlight”) are used by listeners in lexical disambiguation.

Nevertheless, the comparison between stops and fricatives suggests that the acoustic differences between word-initial and word-final /s/ are likely to be less perceptually discernible than those in stops. This may explain why the priming effect in Experiment 2b was stronger when the domain-initial consonants were stops than when they were fricatives. The present study is the first to show that the asymmetry in the production of stops versus fricatives is also reflected in perception. Some sounds, such as stops, may have more articulatory degrees of freedom by nature, whereas other sounds, such as the fricative /s/, may have intrinsically fewer degrees of freedom. This intrinsic difference can mean that some segments may carry more informational

value in speech recognition than others: Stops appear to tell the listener more about their position in the prosodic hierarchy than fricatives do.

#### 4.4. Models of speech production and speech perception

The fundamental assumptions underlying this study are that speakers use prosodic structure as they encode their utterances, and that listeners make use of the acoustic correlates of prosodic structure in speech comprehension. Indeed, we found evidence in support of these assumptions both in production (as reflected in the fine-grained, yet systematic phonetic patterns that we observed in the acoustic analyses of the experimental materials) and in perception (as reflected in the domain-initial strengthening and boundary effects that we observed in the priming experiments). But how do speakers achieve these fine-grained phonetic outcomes, and how do listeners use them?

##### 4.4.1. Production

How does prosodic structure influence the detailed operation of particular articulators? This question concerns the interface between psycholinguistic theories of lexical access in speech production and accounts of articulatory control. The theory of spoken word production developed by Levelt and colleagues (Levelt, 1989; Levelt, Roelofs, & Meyer, 1999; Roelofs, 1997) provides a framework in which to address this question. According to this theory, word form encoding involves spelling out, in abstract phonological terms, various aspects of the selected word form (e.g., its segmental make-up, its syllable structure, and its metrical structure), and adjusting those phonological specifications as a function of the context (“prosodification”, e.g., resyllabification and stress shift). The output of this phonological encoding stage is then fed to the phonetic encoding stage, where the articulatory plan for the word is formed. Phonetic encoding involves, for frequently recurring syllables, retrieval of preprogrammed gestural scores from the “mental syllabary”, or, for infrequent syllables, segment-by-segment assembly of gestural scores (see Browman & Goldstein, 1992, for more information on gestural scores). Fine-tuning of these gestural scores then occurs, as a function of the phonetic context in which each syllable appears (e.g., to account for syllable-to-syllable coarticulation). Finally, the articulatory–motor system converts these gestural plans into motor commands for the different articulators in the vocal tract.

As Levelt et al. (1999) admit, their theory has little to say about the stages of processing which follow syllabary access. But it is precisely at these stages that prosodic structure must influence articulatory planning. In the Levelt et al. theory this is because there is only one stored set of gestural scores per syllable. Irrespective of this theory-internal constraint, however, the locus of the prosodic influence on planning must be post-lexical because the finer-grained phonetic details of individual segments in continuous speech can only be generated after post-lexical processes such as resyllabification and encliticization have specified the surface prosodic structure of the utterance.

Prosodic phenomena such as domain-initial strengthening must therefore arise from an influence of the utterance’s prosodic structure on articulatory planning that occurs after lexical access. Levelt (1989) suggests how this might be achieved, based on a “Prosody Generator”. This device would operate, in the context of the Levelt model, at the interface between phonological and phonetic encoding. It would receive abstract phonological input about selected word-forms (i.e., the output of word-form encoding), and also phrase-structural and pitch-accent information, as well as “intonational meaning” input (i.e., information about the speaker’s emotional state, attitude and rhetorical intentions). Note therefore that abstract prosodic structure in this theory is assumed to be built in parallel to lexical and segmental encoding. It is controversial at which planning stage in the production process higher-level prosodic structure of an utterance is constructed, however. In the Levelt et al. model, higher-level prosodic structure is constructed incrementally as each phonological word is built up in serial order, whereas in the proposal made in Keating and Shattuck-Hufnagel (2002), its construction occurs before lexical selection, that is, in parallel with or immediately after syntactic construction (see also Ferreira, 1993). Whichever way prosodic structure may be constructed, Levelt et al.’s Prosody Generator would be responsible for the syllabification of phoneme strings within phonological phrases (i.e., “prosodification”), and for the specification of the phonetic detail associated with prosodic structure, that is, the rhythm and pitch contour of an utterance, as coded in the duration, loudness and pitch

movement of each syllable. The output of the Prosody Generator therefore includes not only the specification of each syllable in terms of its CV structure and segmental content but also parameters for its duration, amplitude and pitch. Syllabary access retrieves the abstract gestural scores for each (high-frequency) syllable, and phonetic encoding then proceeds to fine-tune those gestural commands according to the prosodic parameters and according to context (i.e., coarticulatory effects).

We suggest that a mechanism akin to the Prosody Generator (i.e., a post-lexical device which integrates the segmental and supra-segmental components of an utterance in order to complete phonological encoding) is a necessary component in any model of speech production. In addition, there must also be a mechanism which translates abstract prosodic specifications into commands for particular articulators. In the [Levelt \(1989\)](#) model, this latter mechanism is separate from the Prosody Generator (it acts after syllabary access, unlike the Prosody Generator). In a model without a syllabary, however, these two mechanisms could interact.

The implementation of prosodically driven phonetic details must ultimately be based on specifications for individual articulators. But parameters such as loudness and duration are not articulator specific. How can these abstract parameters influence all of the articulators involved in the production of any particular segment? For example, how can the command to strengthen the articulation of any given IP-initial consonant be implemented? There are a number of ways by which this might be achieved. [Byrd and Saltzman \(Byrd, 2000; Byrd et al., 2000; Byrd & Saltzman, 2003; Saltzman, 1995\)](#) propose abstract prosodic boundary gestures, specified in the context of a mass–spring dynamical system ([Saltzman & Munhall, 1989](#)). These so-called “ $\pi$ -gestures” act at prosodic boundaries to adjust the temporal extent of all concurrently active constriction gestures (i.e., they temporally shrink or stretch those constriction gestures).  $\pi$ -gestures are hypothesized to modulate the rate of the clock that controls the time-course of gestural activation.

An alternative account is based on the work of [Browman and Goldstein \(1992, 2000\)](#). They proposed that the variation which occurs between gestures can be explained if it is assumed that every gestural phase relation is associated with a differential “bonding strength” (i.e., the degree of cohesion among gestures; see [Cho, 2001](#)). Although this notion was originally motivated to account for the variability in context-sensitive phasing relationships between consonantal gestures, it could be extended to other phasing relationships. Greater bonding strength is hypothesized to trigger stronger gestural cohesion and coarticulatory influence between adjacent gestures. If bonding strength varied, in inverse proportion, to prosodic boundary strength, some phonetic consequences of prosodic structure could be explained. For example, a domain-initial consonant would be less coarticulated than a domain-medial consonant, and could therefore be longer and/or louder.

A third way of accounting for prosodically driven phonetic detail in speech production derives from [Keating’s \(1990\)](#) window model (see also [Byrd, 1996b; Cho, 2002, 2004; Cohn, 1990; Docherty, 1992; Keating, 1996](#); and see, e.g., [Guenther, 1995; Guenther, Hampson, & Johnson, 1998](#), for an independent development of a window model). The targets of articulatory movements are assumed to vary within specified ranges (i.e., windows) rather than simply to have fixed values. [Keating and Shattuck-Hufnagel \(2002\)](#) suggested that the window model could capture prosodically driven phonetic variation: The ranges of the articulations local to a prosodic boundary could be expanded or contracted. [Cho \(2002, 2004\)](#), however, has proposed an alternative formulation in which the total range of variability of a given articulation remains in a fixed window, but in which prosodic factors such as accent and boundary could specify a target region within that window. The target for an articulator associated with a domain-initial consonant could thus be constrained to a more extreme, and narrower range than that for the same articulator, associated with the same consonant, but where that consonant is in domain-medial position.

A radically different type of production model is the exemplar-based approach that has recently been sketched by [Pierrehumbert \(2001, 2002, 2003\)](#). There are two key components in Pierrehumbert’s proposal. First, it is assumed that speech perception involves the storage of exemplars of specific speech events in a multidimensional phonetic space. Phonetic categories are labels defined over “clouds” in different regions of that space, where each member of a cloud is a remembered instance of a given label. Each category is thus associated with a frequency-weighted distribution of phonetic events. The second key assumption is that these perception-based exemplars are used in speech production. Once a phonetic category label has been chosen in production, selection of the motor commands for that label is based on random sampling from the distribution of exemplars associated with that label. It might be possible to extend this class of production model so that it could account for the influence of prosodic structure on phonetic fine detail. The exemplar

clouds could be co-indexed by prosodic domain (e.g., the cloud for [t] could be sub-divided into clouds for IP initial [t], Wd initial [t], and so on). In production, therefore, once the category label and its position in the prosodic hierarchy has been selected, specification of motor commands for that label could be selected at random from the prosodically appropriate sub-distribution of exemplars associated with that label.

The viability of exemplar-based models of prosodically conditioned speech production can only be evaluated properly after the development of more fully specified models. This development will face a number of challenges, however. First, it needs to be established whether the exemplar space is defined in auditory or articulatory terms (Pierrehumbert, 2002, for example, uses both: compare Figs. 1 and 2). If the space is auditory, then there must also be specification in the model of how an auditory exemplar of a category (e.g., a specific instance of IP initial [t]) is linked to the specific articulatory commands required to generate a similar-sounding exemplar. The assumption that the space is instead defined in terms of articulatory dimensions entails the claim that speech perception is based on gestural representations, as for example in Liberman & Mattingly's (1985) motor theory. This seems to run counter to Pierrehumbert's (2003) claims that language acquisition is based primarily on bottom-up statistical analysis of acoustic variability in the infant's ambient language. Second, the size of the representations in the exemplar space will have to be defined more clearly (Pierrehumbert, 2002, focuses on word-sized representations, while Pierrehumbert, 2003, focuses on segments). It is of course possible that both segmental and lexical labels could be associated with the same multidimensional exemplar space. Our argument with respect to prosodic effects in production, however, has been that the specification of prosodically determined articulatory commands must be modulated post-lexically. In addition, several lines of research suggest that spoken-word recognition involves the pre-lexical extraction of abstract phonological information and the subsequent mapping of this information onto abstract lexical representations, contrary to strictly episodic theories of lexical representation (McQueen, Cutler, & Norris, *in press*; Nearey, 1990; Pallier, Colomé, & Sebastián-Gallés, 2001; Slowiaczek, McQueen, Soltano, & Lynch, 2000). Furthermore, an adequate model must explain how speakers can utter novel words (or novel sequences of known words) with the appropriate prosodic characteristics. This would be difficult to achieve if only word-sized episodic representations were available to bias articulation (i.e., it would require specification of an analogical mechanism that would make the correct generalizations over existing lexical episodes). These arguments all suggest that if prosodic effects in production are due to the retrieval of perceptual episodes, then those episodes should be segment-sized rather than word-sized.

There are therefore a number of alternative mechanisms by which an abstract specification of prosodic structure could modulate the motor commands for the particular segment that is required in a given prosodic position. It will be important to distinguish among these alternatives, and to specify in greater detail how the phonetic implementation of prosodic detail is linked to more abstract stages of segmental and supra-segmental encoding. However, although it is controversial at what stage in the production process abstract prosodic structure is constructed, it is already clear that the abstract components of prosodic planning must modulate the process after word form encoding, that is, at the stage of phonetic encoding, so that articulation is tuned accordingly.

#### 4.4.2. Perception

How do listeners use the fine-grained phonetic details that arise at prosodic boundaries? We have argued that this information is used to assist in the segmentation of continuous speech. Most researchers agree that a core component of the spoken-word recognition process is some form of competition among the candidate words which are consistent with the current speech input. As Norris et al. (1997) argued, this competition process can be modulated by a variety of information sources, all of which provide cues to the location of likely word boundaries in the speech signal (e.g., metrical and phonotactic cues to word boundaries). We propose that the acoustic correlates of domain-initial strengthening are another source of word-boundary information (as indeed are the acoustic correlates of domain-final lengthening).

We suggest that, during speech recognition, prosodic structure is built in parallel to lexical/segmental analysis (cf. Christophe et al., 2004; Salverda et al., 2003). In a mirror image to what we have suggested for speech production, where segmental and supra-segmental components are combined late in the speech encoding process (i.e., post-lexically), we propose that parallel analysis of segmental and supra-segmental structure begins early in the speech decoding process (i.e., pre-lexically). Information in the speech signal



which is relevant for phonemic contrasts is used to specify a segmental representation of that signal, for example, in terms of phonemic representations in the Shortlist model (Norris, 1994) and the TRACE model (McClelland & Elman, 1986). In parallel, information which specifies supra-segmental aspects of the signal could be used by a “Prosody Analyzer” to compute the prosodic structure of the current utterance. Some of those computations would likely depend on segmental analysis (e.g., determination of durational cues to prosodic structure could depend on knowledge about the intrinsic duration of specific segments). The proposal is therefore not that the Prosody Analyzer is completely independent of the segmental analysis process; it is that prosodic representations are extracted in parallel to segmental representations. The locations of boundaries in the prosodic structure (i.e., Wd boundaries, Phonological Phrase boundaries and IP boundaries) could then be used to modulate the lexical competition process (which is otherwise driven by the segmental information in the signal). In this proposal, therefore, segmental analysis determines the content in the current input (what words are considered), while the Prosody Analyzer indicates where words are likely to begin and end.

An alternative account is one based on an episodic view of word recognition (e.g., Goldinger, 1996, 1998; Johnson, 1997), which is of course related to the exemplar theory of speech production discussed above. In episodic models of lexical representation, fine-grained phonetic details, including presumably those that arise from domain-initial strengthening, are stored in multiple exemplars of each word. Lexical representations could include the strengthening patterns associated with the onset /t/ in *tickets*, for example, so that speech input manifesting those strengthening patterns would provide a better match to the onset of *tickets* than to the coda of *bust*, its competitor in the sequence *bus tickets*. Prosodic phenomena on an episodic account would therefore influence spoken-word recognition without a specifically prosodic analysis in the comprehension system.

Although the present data are consistent with episodic accounts, there are two reasons to prefer the account which includes a Prosody Analyzer. First, according to our preferred account, a prosodic analysis is made available for other aspects of speech comprehension, such as the assignment of syntactic structure to an utterance. Since listeners appear to compute the prosodic structure of an utterance on-line for syntactic processing (e.g., Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999; Schafer, 1997), they can exploit the same structure in lexical access (cf. Christophe et al., 2004; Salverda et al., 2003). Second, as mentioned in the previous section, some experimental results support the view that abstract phonological information is first extracted, and is subsequently mapped onto abstract lexical representations, counter to strictly episodic theories of the lexicon (McQueen et al., in press; Nearey, 1990; Pallier et al., 2001; Slowiczek et al., 2000).

If the Prosody Analyzer account is the correct one, how then does boundary information influence lexical competition? One possibility is that it does so via the operation of the Possible Word Constraint (PWC; Norris et al., 1997). According to the PWC, sequences of segments in the speech stream which do not include vowels are not possible words in English, and thus parses of the speech stream which contain such sequences are not plausible segmentations. In the implementation of the PWC in Shortlist, therefore, words which are part of parses that include vowelless sequences are penalized in the competition process. Specifically, if there is only consonantal material between the edge of a candidate word (i.e., its onset or its offset) and a likely word boundary, the activation of that candidate is halved. Norris et al. suggest that likely word boundaries can be signaled to the PWC mechanism by metrical and phonotactic cues, by silence, and possibly also by allophonic and acoustic cues. Domain-initial strengthening could provide one such type of word-boundary cue, via the operation of the Prosody Analyzer. In the sequence *bus tickets*, for example, the acoustic correlates of domain-initial strengthening in the [tɪ] could signal to the Prosody Analyzer that there is a possible IP boundary before the [t]. The competitor word *bust* is misaligned with this boundary because there is a single consonant (the [t]) between that boundary and the end of *bust*. This competitor would therefore receive the PWC penalty, allowing *bus* to more easily win the competition process than in the situation where there were no IP-initial strengthening cues.

There are two important points to note about this proposal. First, all of the mechanisms involved must be probabilistic rather than deterministic. For example, the acoustic consequences of domain-initial strengthening found here (e.g., greater amplitude of the burst at the release of the stop) are not absolute indicators that the consonant (e.g., /t/ in *bus # tickets*) is word initial and not word final, nor that there must be an IP boundary before the consonant (indeed, all of the other evidence in the Wd-boundary carrier

sentences in Experiment 2b suggested that there was no IP boundary before, e.g., tickets). Nevertheless, the strengthening cues act to increase the probability of an IP boundary before the consonant, which in turn (via the PWC) decreases the probability that *bust* was the word intended by the speaker. Similarly, parallel extraction of segmental material must also be probabilistic. A large body of evidence (reviewed in McQueen, 2005) suggests that fine-grained segmental detail influences lexical level processes. This suggests that pre-lexical segmental analysis does not involve categorical decisions about abstract sub-lexical representations (e.g., phonemes), but rather that pre-lexical processing is non-deterministic, with probabilistic information flowing in cascade up to the lexical level.

Second, the account based on the PWC penalty cannot be the complete story. Several of the competitor words in the present materials were not misaligned with the prosodic boundary in the way *bust* was (e.g., *fancy* in *fan sites*, where there is a vowel, as well as a consonant, between the end of the competitor and the prosodic boundary). Although there is not a sufficient number of such items to analyze their recognition separately, it is reasonable to suppose that there was a prosodic boundary effect for them. This is because Salverda et al. (2003) have shown prosodic boundary effects in a set of items in which the competitors, in PWC terms, were all aligned with the prosodic boundary. Salverda et al. found that competitors such as *ham* in *hamster* were more strongly activated when the first syllable of *hamster* was cross-spliced from a matched recording where the speaker intended the word *ham* (and thus tended to be longer in duration) than when the syllable came from another recording of *hamster*. They argued that syllable duration provided a cue to a Wd boundary after the monosyllabic word. Neither *ham* nor *hamster* is misaligned with this likely word boundary, however. It is therefore possible that candidate words which are aligned with a prosodic boundary (e.g., *ham* in the cross-spliced *hamster*; *sites* in the cross-spliced *fan sites* in Experiment 2b) receive a boost in activation, making them stronger competitors. It will be important to establish the relative roles of activation-boost mechanisms, such as this, and activation-penalty mechanisms, such as the PWC, in segmentation.

It will also be important to establish whether segmental and prosodic-structure information are extracted separately during word recognition, as assumed in the Prosody Analyzer account. The present results are consistent with a model in which a single mechanism processes both segmental and prosodic details. We prefer the Prosody Analyzer explanation because it makes explicit that segmental and prosodic information have different roles (what vs. where) to play in the word recognition process, and because a prosodic structure is also required for syntactic processing. Whatever the mechanism(s) underlying word recognition may be, however, it is clear that they depend on a device which specifies, on the basis of fine-grained acoustic detail, the location of likely prosodic boundaries in the continuous speech signal.

## 5. Conclusion

This study suggests that the fine-grained phonetic manifestation of prosodic boundaries aids listeners in lexical segmentation. We observed this effect under conditions where we limited the constraints on segmentation imposed by other sources of information. Specifically, we found an effect of domain-initial strengthening where lexical information left an ambiguity about the placement of a word boundary, and where other prosodic cues to boundary location were weaker. We assume that if listeners use the acoustic correlates of domain-initial strengthening to assist segmentation under these conditions, they also use them in other conditions, in parallel with all other available information sources. Our findings therefore suggest that domain-initial strengthening is one of many acoustic cues used in the segmentation of continuous speech. The findings also support our more general hypothesis that speakers signal prosodic structure via systematic phonetic details, and listeners use these cues to prosodic structure in decoding continuous speech.

## Acknowledgment

The authors thank Mark Pitt and Mary Beckman for making testing space available to us at OSU, and Merrill Garrett for use of his lab facilities at the University of Arizona. We also thank two anonymous reviewers for their very constructive comments on a previous version of this paper. Part of this research project was presented at the 15th International Congress of Phonetic Sciences, Barcelona, Spain (McQueen & Cho, 2003).

## Appendix A. Experimental materials

The first line of each item contains a sequence of two critical words. The second words of these sequences were the targets in Experiment 1; the first words were the targets in Experiment 2. In parenthesis next to the two-word sequence is a potential competitor word that straddles the intended word boundary. Each pair of sentences in each item consists of one with a Wd boundary between the two critical words, and one with an IP boundary between the two words. These sentences were used as the auditory prime sentences as well as for splicing (see text for details).

### Stops initial in the second word

- 1 bus tickets (bust)  
John bought several **bus tickets** for his family. (Wd)  
When you get on the **bus**, tickets should be shown to the driver. (IP)
- 2 bow tie (boat)  
The boy will be wearing the **bow tie** to the party. (Wd)  
When you make a **bow**, tie the ribbon tightly. (IP)
- 3 sue poor (soup)  
The bill collection company generally tries not to **sue poor** people. (Wd)  
Since the bill collection company decided to **sue**, poor Dad had to look for free legal advice. (IP)
- 4 day time (date)  
During the **day time** in the summer, we usually work in the garden. (Wd)  
If you are having a tough **day**, time out for a while. (IP)
- 5 pine table (pint)  
Dad made several **pine tables** during the holidays. (Wd)  
Because the carpenters had so much **pine**, tables could be made by them very cheaply. (IP)
- 6 high tide (height)  
I went to see the **high tide** down at the beach. (Wd)  
When your income level goes up **high**, tide over your poorer relatives. (IP)
- 7 chess tournament (chest)  
I participated in several **chess tournaments** last year. (Wd)  
Although he is very good at **chess**, tournaments are always very tough for him. (IP)
- 8 tan tiles (tantilize)  
The couple chose **tan tiles** for their bathroom. (Wd)  
Although they weren't really fond of **tan**, tiles in any other color would have looked terrible in their bathroom. (IP)
- 9 loyal team (loyalty)  
I met several **loyal team** supporters before the Buckeyes game. (Wd)  
When players are **loyal**, team playing skills can be improved. (IP)
- 10 tree top (treat)  
The children climbed up to the **tree top** in order to see the parade. (Wd)  
If the soldiers take a nap under the **tree**, top ranking officers will be mad at them. (IP)
- 11 bee competition (beak)  
The boy participated in a spelling-**bee competition** last year. (Wd)  
If you are not well prepared for the spelling **bee**, competition with last year's winner be useless. (IP)
- 12 mill company (milk)  
To learn about wood products, they visited a **mill company** in Alabama last summer. (Wd)  
When I was thinking about buying a coffee **mill**, company names were the most important things I considered. (IP)
- 13 pie cooking (pike)  
My dad read some good tips about **pie cooking** in the recipe book. (Wd)  
Although it may look easy to bake a **pie**, cooking them well requires a lot of patience. (IP)

- 14 dim parking (dimp)  
The accident happened in the **dim parking** lot when my uncle was visiting us. (Wd)  
When the light is so **dim, parking** can be very difficult. (IP)
- 15 dumb people (dump)  
Believe it or not, I met a whole bunch of **dumb people** at the scientific conference. (Wd)  
Although it makes them look **dumb, people** sometimes wear silly hats. (IP)
- 16 lamb pork (lamp)  
The restaurant is famous for **lamb pork** combo's as their lunch special. (Wd)  
If you don't want to eat **lamb, pork** is the only other alternative. (IP)
- 17 limb pain (limp)  
My mom has been suffering from **limb pain** since last year. (Wd)  
Because the veteran had problems with his artificial **limb, pain** treatment became necessary. (IP)
- 18 bay tourists (bait)  
There will be more **bay tourists** during the summer. (Wd)  
Because of the horrible traffic to the **bay, tourists** generally avoid rush hours to get there. (IP)

#### Fricative /s/ initial in the second word

- 19 ham salad (hamster)  
The boy ate the **ham salad** in the kitchen. (Wd)  
Although the boy wanted some **ham, salad** was the only thing available. (IP)
- 20 neck sizes (next)  
The shirts varied in **neck sizes** to a certain extent. (Wd)  
For a person like you with a very big **neck, sizes** of shirts should be either double or triple X. (IP)
- 21 ten scientists (tense)  
Our institute is recruiting **ten scientists** for a big project. (Wd)  
If you solve problem number **ten, scientists** will want to interview you. (IP)
- 22 ray signals (race)  
The engineer developed several kinds of **ray signals** to perform the analysis. (Wd)  
Although the astronomers attempted to analyze the cosmic **ray, signals** from it were difficult to detect. (IP)
- 23 win summer (wince)  
My brother uses his credit card a lot in June, hoping to **win summer** reductions from the card company. (Wd)  
If you really want to make a big **win, summer** prize draws may be your best bet. (IP)
- 24 whole sales (holster)  
I worked in the **whole sales** department for the company. (Wd)  
On the **whole, sales** pitches should be well prepared to convince picky clients. (IP)
- 25 den search (dense)  
When we found the lions, the **den search** could be called off. (Wd)  
When some of the explorers were found dead in the lion's **den, search** and rescue teams were dispatched to the area. (IP)
- 26 pier side (pierce)  
I got on the Santa Monica bus from the **pier side** to go downtown.  
When you eat at the restaurant near the Santa Monica **Pier, side** orders come free with main dishes. (IP)
- 27 fan sites (fancy)  
The internet company opened new **fan sites** for Harry Potter. (Wd)  
When the movie star started to date her **fan, sites** on the internet covering the latest news sprang up. (IP)
- 28 knee surgery (niece)  
My grandmother received major **knee surgery** last month. (Wd)  
If you have an unbearably painful **knee, surgery** may be the only solution. (IP)
- 29 lock systems (lox)  
They designed quite complicated **lock systems** for the new hotel. (Wd)  
When the company designed a new **lock, systems** for manufacturing it were taken into account. (IP)

- 30 buy salmon (bicycle)  
I went to the store to **buy salmon** for dinner. (Wd)  
Although fish was not what I wanted to **buy, salmon** and cod were the only things that I ended up buying. (IP)
- 31 sea serpents (cease)  
The scientific magazine covered the story about **sea serpents** in this month's issue. (Wd)  
In the old story about a haunted village near the **sea, serpents** were depicted as devils. (IP)
- 32 pick some (pixel)  
The teacher needed to **pick some** students as classroom representatives. (Wd)  
When it was John's turn to take his **pick, some** girl yelled at him. (IP)
- 33 new socks (nuisance)  
On Christmas eve, the little children hang their **new socks** on the hook near the chimney. (Wd)  
Did you know that when shoes are **new, socks** can get worn out more quickly. (IP)
- 34 pen samples (pencil)  
The stationary company sent new **pen samples** to the university store. (Wd)  
Immediately after the stationary company decided to design a new fountain **pen, samples** of other companies' products were examined. (IP)
- 35 pea soup (peace)  
The strange family frequently had **pea soup** for breakfast. (Wd)  
Although the tribe cultivated the **pea, soup** on their island was usually made with beans. (IP)
- 36 fee service (feast)  
The travel agent offered a no **fee service** for bookings made in September. (Wd)  
Although you often have to pay an exorbitant **fee, service** provided by law firms is usually excellent. (IP)
- Nasal /n/ initial in the second word**
- 37 pay notice (pain)  
My employer sent me a **pay notice** every month. (Wd)  
If you are not satisfied with your current **pay, notice** that a salary increase is planned for next year. (IP)
- 38 low net (loan)  
The workers were not happy about the **low net** income they were going to receive. (Wd)  
If the company's turnover is **low, net** profits will also be low. (IP)
- 39 stay nervous (stain)  
If you frighten the children, they will **stay nervous** for several days. (Wd)  
If asked to prolong their **stay, nervous** travelers may become even more anxious. (IP)
- 40 shy nuns (shine)  
I met several **shy nuns** at the church. (Wd)  
Since they are often very **shy, nuns** don't like to talk to strangers. (IP)
- 41 kid near (kidney)  
I didn't see your **kid near** the playground. (Wd)  
If you are looking for my **kid, near** the playground is a good place to look. (IP)
- 42 part names (partner)  
Do you remember the **part names** for the engine parts that have to be fixed? (Wd)  
When I learned my **part, names** for some of the other characters hadn't been chosen yet. (IP)
- 43 play new (plain)  
The actors will **play new** roles in the revised production. (Wd)  
When the author re-wrote the **play, new** roles were created for two lead actors. (IP)
- 44 spy novels (spine)  
The children read several **spy novels** on the train. (Wd)  
After the trial of the British **spy, novels** about espionage became popular. (IP)
- 45 key notions (keen)  
The anthropologist was explaining the **key notions** of her theory at the symposium. (Wd)  
When I lost my **key, notions** about who had taken it came to mind. (IP)

- 46 law notes (lawn)  
The business-studies student lost his **law notes** last week. (Wd)  
During the boring lectures on constitutional **law**, **notes** were taken only by the sharpest students. (IP)
- 47 say negative (sane)  
Christina will probably **say negative** things about our plan. (Wd)  
If ‘no’ is the only answer you can **say**, **negative** feedback is likely what you will receive. (IP)
- 48 lay numerous (lane)  
The farmer expects his hens to **lay numerous** eggs this year. (Wd)  
When the new homeowner shopped for a carpet to **lay**, **numerous** options were available on the market. (IP)

## Appendix B. Acoustic analyses

This table reports the results of a series of paired t-tests between the CVs used for the same-spliced (S-spliced) materials and those used for the cross-spliced (X-spliced) materials. IP-boundary sentences were used as the carrier sentences in the first part of each experiment (Experiments 1a and 2a) and Wd-boundary sentences as the carriers in the second part of each experiment (Experiments 1b and 2b). The spliced versions in each case were created independently using different spoken sentences, so that the same- and cross-spliced CVs differed between those used in IP-boundary carrier sentences (left columns) and those used in Wd-boundary carrier sentences (right columns). The differences (+/–) are the measurement values from IP CV tokens minus the values in Wd CV tokens. Positive values thus indicate the acoustic consequences of IP-initial strengthening in duration (ms) and peak amplitude (dB) measures. \* refers to  $p < 0.05$ ; \*\* to  $p < 0.01$ .

Consonant type	Measures	IP-boundary sentences		Wd-boundary sentences	
		S-spliced (IP)	X-spliced (Wd)	X-spliced (IP)	S-spliced (Wd)
All consonants collapsed	CV duration (across C's)	+ 31.02* (ms)	t(94) = 3.139, p < 0.025	+ 22.88* (ms)	t(94) = 2.258, p < 0.03
	V duration	+ 5.59 (ms)	t(94) = 0.765, p > 0.1	+ 2.69 (ms)	t(94) = 0.363, p > 0.1
	Peak amplitude during V	+ 2.350** (dB)	t(94) = 5.020, p < 0.001	+ 2.347** (dB)	t(94) = 4.976, p < 0.001
/p,t,k/	CV duration (across C's)	+ 10.87 (ms)	t(34) = 0.740, p > 0.1	+ 3.5 (ms)	t(34) = 0.234, p > 0.1
	VOT for stops	+ 3.03 (ms)	t(34) = 0.844, p > 0.1	+ 3.26 (ms)	t(34) = 0.896, p > 0.1
	V duration: stops	+ 7.84 (ms)	t(34) = 0.546, p > 0.1	+ 0.39 (ms)	t(34) = 0.026, p > 0.1
	Peak amplitude during stop burst	+ 3.152** (dB)	t(34) = 3.243, p < 0.01	+ 2.723** (dB)	t(34) = 2.928, p < 0.005
	Peak amplitude during V after stops	+ 2.662** (dB)	t(34) = 3.273, p < 0.005	+ 2.491** (dB)	t(34) = 3.001, p < 0.005
/s/	CV duration: /s/	+ 56.71** (ms)	t(34) = 4.377, p < 0.001	+ 49.12** (ms)	t(34) = 3.820, p < 0.001
	/s/ duration	+ 48.27** (ms)	t(34) = 6.772, p < 0.001	+ 38.29** (ms)	t(34) = 6.538, p < 0.001

	V duration: /s/	+ 8.44 (ms)	t(34) = 0.755, p > 0.1	+ 10.38 (ms)	t(34) = 0.981, p > 0.1
	Peak amplitude during /s/	+ 1.701 (dB)	t(34) = 2.032, p = 0.05	+ 0.995 (dB)	t(34) = 1.197, p > 0.1
	Peak amplitude during V after /s/	+ 2.773** (dB)	t(34) = 3.914, p < 0.005	+ 2.744** (dB)	t(34) = 3.814, p < 0.005
/n/	CV duration: /n/	+ 22.672 (ms)	t(22) = 1.848, p > 0.05	+ 9.98 (ms)	t(22) = 0.935, p > 0.1
	/n/ duration	+ 24.73** (ms)	t(22) = 2.954, p < 0.01	+ 16.06* (ms)	t(22) = 2.206, p < 0.04
	V duration: /n/	−2.057 (ms)	t(22) = 0.189, p > 0.1	−6.068 (ms)	t(22) = 0.588, p > 0.1
	Peak amplitude during /n/	+ 1.056 (dB)	t(22) = 0.975, p > 0.1	+ 0.830 (dB)	t(22) = 0.842, p > 0.1
	Peak amplitude during V after /n/	+ 1.248 (dB)	t(22) = 1.563, p > 0.1	+ 1.533 (dB)	t(22) = 1.831, p > 0.05

### Appendix C. Mean RTs (in milliseconds) and Error Rates (proportion of incorrect responses) when the domain-initial consonant was /n/

Splice condition	Mean RTs (standard errors)				Mean Error Rates (standard errors)			
	S-spliced		X-spliced		S-spliced		X-spliced	
Relatedness	Identical	Unrelated	Identical	Unrelated	Identical	Unrelated	Identical	Unrelated
Experiment 1a (IP sentences)	608 (12.6)	639 (15.1)	600 (14.5)	646 (17.5)	0.02 (0.012)	0.07 (0.022)	0.03 (0.014)	0.07 (0.019)
Experiment 1b (Wd sentences)	579 (13.8)	580 (14.0)	590 (14.0)	581 (13.6)	0.05 (0.017)	0.07 (0.019)	0.03 (0.014)	0.08 (0.021)
Experiment 2a (IP sentences)	542 (14.3)	571 (12.6)	534 (13.2)	566 (13.2)	0.01 (0.010)	0.04 (0.015)	0.01 (0.007)	0.03 (0.014)
Experiment 2b (Wd sentences)	566 (13.8)	558 (13.1)	560 (12.4)	551 (13.7)	0.02 (0.012)	0.03 (0.013)	0.03 (0.018)	0.01 (0.010)

### References

- Barnes, J. (2002). Positional neutralization: A phonologization approach to typological patterns. Ph.D. dissertation. University of California, Berkeley.
- Beckman, J. N. (1998). Positional faithfulness. Ph.D. dissertation. University of Massachusetts, Amherst.
- Beckman, M. E. (1996). The parsing of prosody. *Language and Cognitive Processes*, 11, 17–67.
- Beckman, M. E., & Elam, G. A. (1997). Guidelines for ToBI labelling, version 3.0. Unpublished Manuscript, Ohio State University.
- Beckman, M. E., Edwards, J., & Fletcher, J. (1992). Prosodic structure and tempo in a sonority model of articulatory dynamics. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 68–86). Cambridge: Cambridge University Press.
- Beckman, M. E., & Pierrehumbert, J. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, 3, 255–309.

- Bolinger, D. (1970). Relative height. In P. Léon, G. Faure, & A. Rigault (Eds.), *Prosodic feature analysis* (pp. 109–127). Montreal: Marcel Didier (Reprinted in Bolinger, D. (1972). *Intonation: Selected readings*. Harmondsworth, Penguin Books.).
- Browman, C., & Goldstein, L. (1992). Articulatory phonology: An overview. *Phonetica*, 49, 155–180.
- Browman, C. P., & Goldstein, L. (2000). Competing constraints on intergestural coordination and self-organization of phonological structures. *Bulletin de la Communication Parlée*, 5, 25–34.
- Byrd, D. (1996a). Influences on articulatory timing in consonant sequences. *Journal of Phonetics*, 24, 209–244.
- Byrd, D. (1996b). A phase window framework for articulatory timing. *Phonology*, 13, 139–169.
- Byrd, D. (2000). Articulatory vowel lengthening and coordination at phrasal junctures. *Phonetica*, 57, 3–16.
- Byrd, D., Kaun, A., Narayanan, S., & Saltzman, E. (2000). Phrasal signatures in articulation. In M. Broe, & J. Pierrehumbert (Eds.), *Acquisition and the lexicon: Papers in laboratory phonology V* (pp. 70–88). Cambridge, UK: Cambridge University Press.
- Byrd, D., & Saltzman, E. (2003). The elastic phrase: Modeling the dynamics of boundary-adjacent lengthening. *Journal of Phonetics*, 31, 149–180.
- Carlson, K., Clifton, C., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, 45, 58–81.
- Cho, T. (2001). Effects of morpheme boundaries on intergestural timing. Evidence from Korean. *Phonetica*, 58, 129–162.
- Cho, T. (2002). *The effects of prosody on articulation in English*. New York, NY: Routledge.
- Cho, T. (2004). Prosodically conditioned strengthening and vowel-to-vowel coarticulation in English. *Journal of Phonetics*, 32, 141–176.
- Cho, T. (2005). Prosodic strengthening and featural enhancement: Evidence from acoustic and articulatory realizations of /a,i/ in English. *Journal of the Acoustical Society of America*, 117(6), 3867–3878.
- Cho, T. (in press). Manifestation of prosodic structure in articulation: Evidence from lip kinematics in English. In L. M. Goldstein, D. H. Whalen, & C. T. Best (Eds.), *Laboratory phonology 8: Varieties of phonological competence*. Berlin & New York: Mouton de Gruyter.
- Cho, T., & Jun, S. (2000). Domain-initial strengthening as featural enhancement: Aerodynamic evidence from Korean. *Chicago Linguistics Society*, 36, 31–44 (Also appeared in *UCLA Working Papers in Phonetics*, 99, 57–70.).
- Cho, T., & Keating, P. (2001). Articulatory and acoustic studies of domain-initial strengthening in Korean. *Journal of Phonetics*, 29, 155–190.
- Choi, H. (2003). Prosody-induced acoustic variation in English stop consonants. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*. Adelaide: Causal Productions.
- Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language*, 51, 523–547.
- Cohn, A. (1990). *Phonetic and phonological rules of nasalization*. UCLA Working Papers in Phonetics (Vol. 76) (Ph.D. dissertation, UCLA).
- Cole, J., Choi, H., Kim, H., & Hasegawa-Johnson, M. (2003). The effect of accent on the acoustic cues to stop voicing in Radio News speech. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*. Adelaide: Causal Productions.
- Cooper, A. (1991). *Glottal gestures and aspiration in English*. Ph.D. dissertation, Yale University.
- Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.
- 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.
- De Jong, K. (1995). The supraglottal articulation of prominence in English: Linguistic stress as localized hyperarticulation. *Journal of the Acoustical Society of America*, 97, 491–504.
- Docherty, G. J. (1992). *The timing of British English obstruents*. Berlin: Foris.
- Dumay, N., Frauenfelder, U. H., & Content, A. (2000). Acoustic–phonetic cues and lexical competition in segmentation of continuous speech. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the workshop on spoken word access processes* (pp. 127–130). Nijmegen: MPI for Psycholinguistics.
- Edwards, J. E., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America*, 89, 369–382.
- Ferreira, F. (1993). Creation of prosody during sentence production. *Psychological Review*, 100, 233–253.
- Fougeron, C. (1999). Prosodically conditioned articulatory variations: A review. *UCLA Working Papers in Phonetics*, 97, 1–74.
- Fougeron, C. (2001). Articulatory properties of initial segments in several prosodic constituents in French. *Journal of Phonetics*, 29, 109–135.
- Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, 101(6), 3728–3740.
- Gaskell, M. G., & Marslen-Wilson, W. D. (1997). Integrating form and meaning: A distributed model of speech perception. *Language and Cognitive Processes*, 12, 613–656.
- Goldinger, S. D. (1996). Words and voices: Episodic traces in spoken word identification and recognition memory. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 22, 1166–1183.
- Goldinger, S. D. (1998). Echoes of echoes?: An episodic theory of lexical access. *Psychological Review*, 105, 251–279.
- Gout, A., Christophe, A., & Morgan, J. L. (2004). Phonological phrase boundaries constrain lexical access: II. Infant data. *Journal of Memory and Language*, 51, 548–567.
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, 45, 133–159.
- Gow, D. W. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, 28, 163–179.



- Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance*, 21, 344–359.
- Gow, D. W., Melvold, J., & Manuel, S. (1996). How word onsets drive lexical access and segmentation: Evidence from acoustics, phonology and processing. In H. T. Bunnell, & W. Idsardi (Eds.), *Proceedings of ICSLP 96 (CD-ROM)*. Philadelphia: University of Delaware & Alfred I. duPont Institute.
- Guenther, F. H. (1995). Speech sound acquisition, coarticulation, and rate effects in a neural network model of speech production. *Psychological Review*, 102, 594–621.
- Guenther, F. H., Hampson, M., & Johnson, D. (1998). A theoretical investigation of reference frames for the planning of speech movements. *Psychological Review*, 105, 611–633.
- Harrington, J. (2003). Commentary: Consonant strengthening and lengthening in various languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology 6: Phonetic interpretations* (pp. 183–196). Cambridge, UK: Cambridge University Press.
- Johnson, E. K. (2003a). Word segmentation during infancy: The role of subphonemic cues to word boundaries. PhD. Dissertation, The Johns Hopkins University.
- Johnson, E. K. (2003b). Speaker intent influences infants' segmentation of potentially ambiguous utterances. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 1995–1998). Adelaide: Causal Productions.
- Johnson, K. (1997). Speech perception without speaker normalization. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–166). San Diego: Academic Press.
- Jun, S.-A. (1993). *The phonetics and phonology of Korean prosody*. Ph.D. dissertation, Ohio State University.
- Jun, S.-A. (1998). The accentual phrase in the Korean prosodic hierarchy. *Phonology*, 15, 189–226.
- Keating, P. A. (1990). The window model of coarticulation: Articulatory evidence. In J. Kingston, & M. Beckman (Eds.), *Papers in laboratory phonology I: Between the grammar and the physics of speech* (pp. 451–470). Cambridge: Cambridge University Press.
- Keating, P. A. (1996). The phonology–phonetics interface. In U. Kleinhenz (Ed.), *Interfaces in phonology* (pp. 262–278). Berlin: Studia grammatica 41, Akademie Verlag.
- Keating, P. A., Cho, T., Fougeron, C., & Hsu, C. (2003). Domain-initial strengthening in four languages. In J. Local, R. Ogden, & R. Temple (Eds.), *Papers in laboratory phonology 6: Phonetic interpretations* (pp. 145–163). Cambridge, UK: Cambridge University Press.
- Keating, P. A., & Shattuck-Hufnagel, S. (2002). A prosodic view of word form encoding for speech production. *UCLA Working Papers in Phonetics*, 101, 112–156.
- Keating, P. A., Wright, R., & Zhang, J. (1999). Word-level asymmetries in consonant articulation. *UCLA Working Papers in Phonetics*, 97, 157–173.
- Kim, S. (2003a). Domain-initial strengthening of Korean fricatives. In S. Kuno, et al. (Eds.), *Harvard studies in Korean linguistics* (Vol. IX, pp. 164–173).
- Kim, S. (2003b). The role of post-lexical tonal contours in word segmentation. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences*. Adelaide: Causal Productions.
- Kirk, C. (2000). Syllabic cues to word segmentation. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the Workshop on Spoken Word Access Processes* (pp. 131–134). Nijmegen: MPI for Psycholinguistics.
- Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and inference in the resolution of temporal syntactic closure ambiguity. *Journal of Memory and Language*, 40, 153–194.
- Klatt, D. H. (1975). Vowel lengthening is syntactically determined in connected discourse. *Journal of Phonetics*, 3, 129–140.
- Ladefoged, P. (2000). *A Course in Phonetics* (4th Ed.). Fort Worth: Harcourt, Brace, Jovanovich.
- Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, 51, 2018–2024.
- Levelt, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge: MIT Press.
- Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Brain and Behavioral Sciences*, 22, 1–38.
- Lieberman, A. M., & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition*, 21, 1–36.
- Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear and Hearing*, 19, 1–36.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, 25, 71–102.
- Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology*, 10, 29–63.
- 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 Publications.
- McQueen, J. M., & Cho, T. (2003). The use of domain-initial strengthening in segmentation of continuous English speech. In M. J. Solé, D. Recasens, & J. Romero (Eds.), *Proceedings of the 15th international congress of phonetic sciences* (pp. 2993–2996). Adelaide: Causal Productions.
- McQueen, J. M., Cutler, A., & Norris, D. (in press). Phonological abstraction in the mental lexicon, *Cognitive Science*.
- Nakatani, L. H., & Dukes, K. D. (1977). Locus of segmental cues for word juncture. *Journal of the Acoustical Society of America*, 62, 714–719.
- Nearey, T. M. (1990). The segment as a unit of speech perception. *Journal of Phonetics*, 18, 347–373.
- Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris Publications.
- Nooteboom, S. G. (1981). Lexical retrieval from fragments of spoken words: Beginnings vs. endings. *Journal of Phonetics*, 9, 407–424.
- 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.
- Pallier, C., Colomé, A., & Sebastián-Gallés, N. (2001). The influence of native-language phonology on lexical access: Exemplar-based versus abstract lexical entries. *Psychological Science*, 12, 445–449.
- Pierrehumbert, J. (1980). *The phonology and phonetics of English intonation*. Ph.D. dissertation, MIT.
- Pierrehumbert, J. (2001). Exemplar dynamics: Word frequency, lenition, and contrast. In J. Bybee, & P. Hopper (Eds.), *Frequency effects and the emergence of linguistic structure* (pp. 137–157). Amsterdam: John Benjamins.
- Pierrehumbert, J. (2002). Word-specific phonetics. In C. Gussenhoven, & N. Warner (Eds.), *Laboratory phonology VII* (pp. 101–140). Berlin: Mouton de Gruyter.
- Pierrehumbert, J. (2003). Phonetic diversity, statistical learning, and acquisition of phonology. *Language and Speech*, 46(2–3), 115–154.
- Pierrehumbert, J., & Beckman, M. E. (1988). *Japanese tone structure*. Cambridge, MA: MIT Press.
- Pierrehumbert, J., & Talkin, D. (1992). Lenition of /h/ and glottal stop. In G. Docherty, & D. R. Ladd (Eds.), *Papers in laboratory phonology II: Gesture, segment, prosody* (pp. 90–117). Cambridge, UK: Cambridge University Press.
- Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, 20, 331–350.
- Recasens, D. (1999). Lingual coarticulation. In W. Hardcastle, & N. Hewlett (Eds.), *Coarticulation: Theory, data and techniques* (pp. 80–104). Cambridge: Cambridge University Press.
- Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition*, 64, 249–284.
- Saltzman, E. (1995). Intergestural timing in speech production: Data and modeling. In K. Elenius, & P. Brandrud (Eds.), *Proceedings of the 13th international congress of phonetic sciences*, Vol. 2 (pp. 84–91). Stockholm: Stockholm University.
- Saltzman, E., Munhall, K.G. (1989). A dynamical approach to gestural patterning in speech production. *Ecological Psychology*, 1, 333–382.
- Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition*, 90, 51–89.
- Schafer, A. J. (1997). *Prosodic parsing: The role of prosody in sentence comprehension*. University of Massachusetts Doctoral Dissertation.
- Selkirk, E. (1984). *Phonology and syntax: The relation between sound and structure*. Cambridge: MIT Press.
- Selkirk, E. (1986). On derived domains in sentence phonology. *Phonology Yearbook*, 3, 371–405.
- Shadle, C. H., & Scully, C. (1995). An articulatory–acoustic–aerodynamic analysis of [s] in VCV sequences. *Journal of Phonetics*, 23, 53–66.
- Shattuck-Hufnagel, S. (1986). The representation of phonological information during speech production planning: Evidence from vowel errors in spontaneous speech. *Phonology Yearbook*, 3, 117–149.
- Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, 25, 193–247.
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, 68, 1–16.
- Silverman, K., Beckman, M., Pitrelli, J., Ostendorf, M., Wightman, C., Price, P., et al. (1992). TOBI: A standard for labeling English prosody. In J. J. Ohala, T. M. Nearey, B. L. Derwing, M. M. Hodge, & G. E. Wiebe (Eds.), *Proceedings of the 1992 international conference on spoken language processing*, vol. 2 (pp. 867–870). Edmonton: University of Alberta.
- Slowiaczek, L. M., McQueen, J. M., Soltano, E. G., & Lynch, M. (2000). Phonological representations in prelexical speech processing: Evidence from form-based priming. *Journal of Memory and Language*, 43, 530–560.
- Smith, R. (2004). *The role of fine phonetic detail in word segmentation*. Ph.D. dissertation, University of Cambridge.
- Smith, R., & Hawkins, S. (2000). Allophonic influences on word-spotting experiments. In A. Cutler, J. M. McQueen, & R. Zondervan (Eds.), *Proceedings of the workshop on spoken word access processes* (pp. 139–142). Nijmegen: MPI for Psycholinguistics.
- Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, 48, 233–254.
- Steriade, D. (1999). Phonetics in phonology: The case of laryngeal neutralization. In: M. Gordon (Ed.), *UCLA working papers in linguistics*, No. 2: *Papers in phonology* (Vol. 3, pp. 25–146).
- Tabain, M. (2003a). Effects of prosodic boundary on /aC/ sequences: Acoustic results. *Journal of the Acoustical Society of America*, 113, 516–531.
- Tabain, M. (2003b). Effects of prosodic boundary on /aC/ sequences: Articulatory results. *Journal of the Acoustical Society of America*, 113, 2834–2849.
- Tabossi, P., Collina, S., Mazzetti, M., & Zoppello, M. (2000). Syllables in the processing of spoken Italian. *Journal of Experimental Psychology: Human Perception and Performance*, 26, 758–775.
- White, L. (2002). *English speech timing: A domain and locus approach*. Ph.D. dissertation, University of Edinburgh.
- Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America*, 91, 1707–1717.