Available online at www.sciencedirect.com
SCIENCE DIRECT®
COGNITION
ELSEVIER
Cognition 90 (2003) 51–89
www.elsevier.com/locate/COGNIT

The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension

Anne Pier Salverda*, Delphine Dahan, James M. McQueen

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Received 8 August 2002; revised 27 January 2003; accepted 20 June 2003

Abstract

Participants' eye movements were monitored as they heard sentences and saw four pictured objects on a computer screen. Participants were instructed to click on the object mentioned in the sentence. There were more transitory fixations to pictures representing monosyllabic words (e.g. *ham*) when the first syllable of the target word (e.g. *hamster*) had been replaced by a recording of the monosyllabic word than when it came from a different recording of the target word. This demonstrates that a phonemically identical sequence can contain cues that modulate its lexical interpretation. This effect was governed by the duration of the sequence, rather than by its origin (i.e. which type of word it came from). The longer the sequence, the more monosyllabic-word interpretations it generated. We argue that cues to lexical-embedding disambiguation, such as segmental lengthening, result from the realization of a prosodic boundary that often but not always follows monosyllabic words, and that lexical candidates whose word boundaries are aligned with prosodic boundaries are favored in the word-recognition process.

Keywords: Spoken-word recognition; Embedded words; Eye movements; Prosodic structure

1. Introduction

A fundamental characteristic of speech is that it extends over time. Spoken words are temporal sequences that become fully available to the listener only after a few hundred milliseconds. A large body of evidence has now established that the recognition of a spoken word proceeds incrementally, as soon as acoustic information becomes available.

* Corresponding author. Present address: Department of Psychology, University of York, Heslington, York, YO10 5DD, UK.
E-mail address: a.salverda@psych.york.ac.uk (A.P. Salverda).

0022-2860/$ - see front matter © 2003 Elsevier B.V. All rights reserved.
doi:10.1016/S0010-0277(03)00139-2

Words that are consistent with the acoustic signal are activated and compete for recognition (e.g. Luce, 1986a; Marslen-Wilson, 1987; McQueen, Norris, & Cutler, 1994; Zwitserlood, 1989). Because partial spoken input is often consistent with multiple lexical interpretations, the recognition of a spoken word can be viewed as a process of ambiguity resolution. For example, the initial sounds of the word *candle*, /kænd/, are also consistent with the word *candy*. Subsequent information disambiguates between alternatives, often allowing words to be recognized before their offset.

However, a large proportion of words cannot be uniquely identified before their offset but only after a portion of the subsequent context has been heard (Bard, Shillcock, & Altmann, 1988; Grosjean, 1985). One reason for such delayed recognition is that many words are embedded at the onset of other, longer words. For example, the phonemic sequence /kæn/ matches the word *can* but also the onset of longer words such as *candy* or *candle*. The attribution of the sequence to a specific lexical item may be delayed, as well as that of the segments following the sequence, if together they phonemically match a long candidate. For example, the phoneme /d/ following the sequence /kæn/ in the phrase *can do* should not be interpreted as providing unambiguous support for the interpretation *candy*. Onset-embedded words therefore present a potentially acute problem for word recognition. The incoming acoustic signal is processed incrementally, but this signal may sometimes be unambiguously attributed to a specific lexical item only after a substantial time delay. The present research addresses how lexical embedding and incrementality in spoken-word recognition can be reconciled. We will argue that the speech signal can contain fine-grained information that listeners use to disambiguate longer words with lexical embeddings from tokens of those shorter, embedded words. Specifically, we will argue that the speech signal contains cues resulting from the realization of prosodic boundaries, and that words that are aligned with such boundaries are favored in the activation and competition process that leads to word recognition.

All current models of spoken-word recognition capture the process of ambiguity resolution during word recognition by assuming some form of competition between simultaneously activated candidates. The mechanism by which competition is instantiated differs across models, depending in part on the models' lexical representations. In some localist connectionist models, such as TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994), word candidates that match the same part of the speech signal compete with each other via inhibitory inter-word connections. Competition is also present in the Distributed Cohort Model (DCM; Gaskell & Marslen-Wilson, 1997, 1999), although competition is a consequence of the model's representations and architecture, rather than an added component. In this model, a simple recurrent network is trained to map input sequences onto a set of features representing the current word. The same set of features encodes patterns associated with any word. Upon partial input, the model generates a blend of the activation patterns associated with all the words that are consistent with the available input. Thus, competition takes the form of interference between the patterns associated with all lexical candidates that are consistent with partial input.

Lexical embedding presents a problem for distributed connectionist models based on a recurrent network because, in these models, the network is trained to activate a representation of the *current* word in a sequence (Elman, 1990; Norris, 1990). An

embedded word can be identified with certainty only once post-offset information is available, but, by the time this information is available, the representation of the following word will already be activated in the network. The model is therefore unable to modify the representations activated by the previous word. Thus, the representation associated with a short word can never be fully activated. Solutions to this problem have been proposed. One consists of training a network to activate representations of word sequences (e.g. Davis, Gaskell, & Marslen-Wilson, 1997). Because the network needs to maintain a representation of all the words in the sequence, it is able to use the following context to identify short words. Another is to consider recognition as a two-stage process (Norris, 1994). At the first stage, a recurrent network could continuously generate (localist) lexical hypotheses. These hypotheses would then enter a second stage, where they compete with one another, on the basis of their degree of support in the input. Short words could be recognized because word candidates would compete not only with other words beginning at the same time, but also with words beginning earlier or later in the signal (i.e. candidates that were selected by the recurrent network during its processing of other portions of the input).

Competition via inter-word inhibition can account for the recognition of short words such as *can* and longer, carrier words such as *candy* (Frauenfelder & Peeters, 1990; McQueen, Cutler, Briscoe, & Norris, 1995; Norris, 1994). All words matching the ambiguous sequence (i.e. the embedded word and its carrier words) remain active candidates until the input is disambiguated. The later in time disambiguating information becomes available, the longer it takes for the ambiguity to be resolved. The disambiguating information can act to penalize the candidates that mismatch it, as in Shortlist, or to boost the activation of other words that compete with the mismatching candidates, as in TRACE and Shortlist. For example, the carrier word *candy* will receive inhibition from the candidates *do* and *doom* (amongst others) when the vowel information /uː/ in the phrase *can do* becomes available, allowing the word *can* to account for the sequence /kæn/. In localist models without inter-word inhibition, a penalty assigned to candidates that mismatch the input will allow the short word to be recognized.

Regardless of how competition is instantiated, lexical embedding appears to impose strong constraints on the recognition of spoken words in continuous speech. It requires that listeners (a) can evaluate lexical parsings that may comprise more than one word (i.e. the activation of representations of sequences of words rather than of a single, current word), and (b) can revise degrees of evidence for a lexical parsing substantially later in the speech stream, when disambiguating information becomes available. Because onset embedding is a prevalent phenomenon in languages (as evaluated from machine-readable dictionaries of English and Dutch; Frauenfelder, 1991; Luce, 1986b; McQueen et al., 1995), these constraints need to be addressed by models of spoken-word recognition.

The lexical ambiguity resulting from onset embedding, as just described, is especially acute if the sequence shared by the short word and the longer, carrier word is fully ambiguous. Thus far, we have assumed that the ambiguous sequence (e.g. /kæn/) is indistinguishable whether it is produced as a monosyllabic word (e.g. *can*) or as the initial portion of a carrier word (e.g. excised from *candle*). However, some factors might contribute to reduce, or even eliminate, the ambiguity. Syllable match is one of them. A monosyllabic word and a carrier word may not be strong competitors if their syllable structures do not match. For example, the sequence /siːl/ is phonemically embedded in

ceiling at onset, but the *l* corresponds to the onset of the second syllable in *ceiling* and to the syllable coda in *seal*. Syllabic structure has robust acoustic consequences on the realization of the segments of the sequence. In the *seal/ceiling* case, for example, the /l/ will change from the dark, coda allophone in *seal* to the light onset allophone in *ceiling* (Abercrombie, 1967; Jones, 1972).

Furthermore, listeners have been shown to use the acoustic cues to syllabic structure that are available in the speech signal to favor the candidate words that match that syllabic structure (Tabossi, Collina, Mazzetti, & Zoppello, 2000). In a study that is more directly related to the problem of lexical embedding, Quené (1992) used ambiguous two-word sequences such as the Dutch phrases *diep in* and *die pin* and showed that Dutch listeners make use of variations in the intervocalic-consonant duration to assign a syllabic structure, and, as is the case in his stimuli, a word boundary. Vroomen and de Gelder (1997) found no evidence for the activation of an embedded word that mismatched the syllabic structure of its carrier word (e.g. the Dutch word *vel* was not activated upon hearing the carrier word *velg*), but did find evidence for the activation of a word embedded in a nonword that mismatched its syllabic structure (e.g. the word *vel* was activated upon hearing the nonword **velk*). This suggests that syllabic mismatch with the input alone does not rule out an embedded candidate.

Even with matched syllabic structure, the ambiguity in assigning a sequence to an embedded word or its carrier word may be reduced by fine-grained acoustic cues present in the sequence itself. This possibility was evaluated in a recent study conducted by Davis, Marslen-Wilson, and Gaskell (2002). They compared the estimated degree of activation of an embedded word (e.g. *cap*) and its carrier word (e.g. *captain*) when listeners were exposed to an ambiguous sequence that originated either from a short word (e.g. /kæp/ from the word *cap*, as in the sentence *the soldier saluted the flag with his cap tucked under his arm*) or from the onset of a matched longer word (e.g. /kæp/ from the word *captain*, as in the sentence *the soldier saluted the flag with his captain looking on*). The ambiguity was maximized by keeping the consonant following the sequence identical in both cases (e.g. *cap* was followed by a word starting with the consonant /t/, i.e. *tucked*). The results suggested that there was differential activation for the shorter and longer words in each version of the sequence, with more activation for the shorter word when the sequence came from a shorter word than when it came from a longer word, and more activation for the longer word when the sequence came from a longer word than when it came from a shorter word. Acoustic analyses of the stimuli indicated systematic differences in the duration of the sequence. The sequence was longer when it was a monosyllabic word (291 ms) than when it corresponded to the initial syllable of a carrier word (243 ms). These durational differences were associated with (less systematic) F0 differences. The mean F0 on the vowels of monosyllabic words tended to be lower than on the vowels of the initial syllables of the longer words. Analyses of the same utterances produced by three additional speakers who were naive to the purpose of the study confirmed the presence of durational and F0 differences in the ambiguous sequence as a function of the word it originated from. Davis et al. concluded that "cues are present in the speech stream that assist the perceptual system in distinguishing short words from the longer competitors in which they are embedded" (Davis et al., 2002, p. 238).

Davis et al.'s (2002) study is important because it constitutes the first demonstration that the ambiguity resulting from onset lexical embedding is not necessarily as severe as

a linear phonemic transcription of the monosyllabic word and its carrier word implies. However, it does not speak to the issue of what may cause the productions of monosyllabic words and initial portions of longer words to differ acoustically, nor how these acoustic cues can differentially contribute to the activation of monosyllabic or longer candidate words. One possibility is to view these acoustic differences as inherent properties of the words themselves, that is, as properties that are specified lexically in the speech-production system. The specification that a monosyllabic word is longer than the corresponding first syllable of a carrier word would be similar to the specification of other between-word differences (e.g. that the /l/ in *seal* is dark but is light in *ceiling)*. These durational characteristics (and perhaps other differences) would be represented as stored knowledge associated with short and long words, which would constrain the phonetic realization of these words in production.

An alternative hypothesis is that acoustic differences between the production of monosyllabic words and the initial portions of longer words are determined by prosodic factors, whose origin is external to the words themselves. Acoustic differences such as durational distinctions between syllables in different types of words would arise as a consequence of production mechanisms that specify the prosodic structure of utterances. A sequence realized as a monosyllabic word would be characterized by acoustic cues favoring a monosyllabic interpretation insofar as the prosodic boundary following the monosyllabic word was phonetically instantiated.

Davis et al. (2002) dismissed the role of prosody in accounting for the duration and F0 differences in their original stimuli. They argued that there was no prosodic boundary after the embedded words in their utterances. The duration differences they reported (and, to some extent, the F0 differences), however, lead us to believe that a prosodic boundary was present, even though its acoustic realization did not involve a silent pause. Segments, especially vowels, tend to be longer in preboundary positions (Klatt, 1976; Lehiste, 1972; Martin, 1970; Oller, 1973, for English; Cambier-Langeveld, 2000; Nooteboom & Doodeman, 1980, for Dutch). Segmental lengthening is strong before an utterance boundary (as in words in isolation), but can also be found at more minor phrase boundaries. The effect of a word boundary on segment durations when the word boundary does not also correspond to a phrase boundary has been viewed as less systematic (e.g. Harris & Umeda, 1974). However, other studies have shown that segments that appear at the edge of a (prosodic) word constituent tend to be longer than segments further from the edge (e.g. Beckman & Edwards, 1990; Turk & Shattuck-Hufnagel, 2000). For example, Turk and Shattuck-Hufnagel (2000) showed that the sequence /tuːn/ is longer in *tune acquire* than in *tuna choir*.

The lengthening of segments in preboundary positions has been integrated into a general framework that aims to account for systematic variations in the production of segments by resorting to the concept of prosodic domain (Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986; see Shattuck-Hufnagel & Turk, 1996, for a review). The prosodic constituents of an utterance are in part determined by the utterance's morphosyntactic structure, so that acoustic correlates to prosodic boundaries mark linguistic constituents (e.g. Cooper & Paccia-Cooper, 1980; but see Pierrehumbert & Liberman, 1982; Shattuck-Hufnagel & Turk, 1996, and references therein, for discussions on the mapping between syntax and prosody). Ladd and Campbell (1991)

and Wightman, Shattuck-Hufnagel, Ostendorf, and Price (1992), amongst others, have shown that the amount of preboundary lengthening varies with the level of the prosodic boundary. Segmental lengthening is stronger at the edge of high prosodic domains, such as intermediate and intonational phrases, than at the edge of lower prosodic domains, such as prosodic words and accentual phrases. This was confirmed in Dutch by Cambier-Langeveld (2000). The prosodic structure of an utterance can also affect segmental articulation. Fougeron and Keating (1997), for example, showed that segments located in the immediate vicinity of the edge of a prosodic domain (in particular, initial consonants and final vowels) have more extreme lingual articulation, a phenomenon they refer to as articulatory strengthening. Because the boundaries of prosodic words, accentual phrases, and any higher prosodic domains are always aligned with a lexical-word boundary, any acoustic cues marking the edge of these prosodic domains could help disambiguate monosyllabic, embedded words from their carrier words before post-offset information is heard.

There is evidence that the acoustic correlates of some prosodic domains, although subtle, are perceptually salient. For instance, Christophe and her colleagues (Christophe, Dupoux, Bertoncini, & Mehler, 1994; Christophe, Mehler, & Sebastián-Gallés, 2001) demonstrated that newborns discriminate bisyllabic sequences as a function of the prosodic environment they originated from (i.e. sequences from within a word or sequences straddling a phonological-phrase boundary, such as the sequence *latí* embedded in the Spanish word *gelatína* or in the phrase *Manuéla tímida*, respectively). Acoustic analyses indicated that duration, F0, and energy of the preboundary vowel varied with the prosodic environment, although not all three parameters always showed systematic differences.

In the present study, we revisited the issue of lexical embedding with this prosodic perspective in mind. We conducted a series of experiments to investigate the conditions under which the production of a monosyllabic or longer word contributes to lexical disambiguation. If listeners' discrimination of an ambiguous sequence as a monosyllabic word or the onset of a longer word depends on the prosodic context in which the sequence was produced, we should expect between- as well as within-sentence variability. As mentioned earlier, the morphosyntactic structure of a sentence imposes constraints on the choices that a speaker makes among the prosodic possibilities for a given sentence. These choices are further influenced by other performance factors, such as speech rate and the length and symmetry of constituent-boundary locations (e.g. Gee & Grosjean, 1983). Thus, the precise prosodic phrasing of a particular sentence can vary widely. The degree to which a monosyllabic word can be discriminated from the initial portion of a longer word should therefore depend on acoustic correlates to prosodic boundaries, such as segmental lengthening. Note that the influence of some prosodic phenomena on lexical disambiguation, such as the presence of a major prosodic boundary after the monosyllabic word (realized in part by the presence of a large, silent pause), is not subject to controversy. Our goal was to evaluate the prosodic modulation of this disambiguation in conditions similar to those used by Davis et al. (2002), that is, in continuous speech with no obvious interruption produced after the monosyllabic word.

We examined the prosodic-boundary hypothesis in two ways. First, the prosodic context in which the monosyllabic word was produced was varied. The monosyllabic word was followed by either a stressed or an unstressed syllable (Experiment 1). A Dutch speaker, naive to the purpose of the experiment, produced Dutch sentences that

contained either a polysyllabic carrier word (e.g. the word *hamster* in *ze dacht dat die hamster verdwenen was*, she thought that that hamster had disappeared) or a monosyllabic word that matched the first syllable of the carrier word (e.g. the word *ham* in *ze dacht dat die ham stukgesneden was*, she thought that that ham had been sliced). The first syllable of the word following the monosyllabic word was either stressed or unstressed (e.g. *ham 'stukgesneden* vs. *ham ste'riel*). The stress status of the syllable following the monosyllabic word was not controlled in the Davis et al. (2002) stimuli, even though it is a potentially important factor. Indeed, the presence of a stressed syllable rather than an unstressed syllable after the (stressed) monosyllabic word may induce the realization of a prosodic juncture after the monosyllabic word because such a boundary would lessen the potential clash between two adjacent stresses. This in turn could affect the realization of the monosyllabic word itself, modulating the degree to which the speech signal could be lexically disambiguated.[1]

Second, we evaluated how systematically the production of monosyllabic or longer words provides disambiguating cues by selecting recorded tokens of each on the basis of their duration (Experiments 2 and 3). As the results will show, the presence of variability in the acoustic realization of those sequences, as well as the impact of this variability on lexical disambiguation, indicate that the lexical interpretation of an embedded sequence is determined by its duration, rather than by its source (i.e. the word it originated from). This is consistent, we will argue, with the hypothesis that the disambiguation of lexical embedding mostly depends on the presence of acoustic cues that mark a prosodic boundary, such as segmental lengthening.

In order to isolate the effect of the realization of the ambiguous sequence from the effect of its following context on lexical interpretation, Davis et al. (2002) presented sentences truncated at different points in the speech signal (i.e. at the offset of the ambiguous sequence, at the onset of the disambiguating phoneme, etc.), and probed activation for the monosyllabic or carrier lexical interpretation at each of these points. Any differential activation observed at each of these points was attributed to the acoustic information presented up to the truncation point. However, as shown by Zwitserlood and Schriefers (1995), sensory information and its impact on lexical activation may not always be tightly time-locked. Attributing effects on lexical activation to a specific part of the speech signal may therefore be difficult.

We took a different approach. We used cross-splicing to evaluate the effect of the realization of the ambiguous sequence on lexical activation. The initial part of the sentence that mentioned the carrier word, up to and including the first syllable of the carrier word (e.g. *ze dacht dat die ham[ster]*, she thought that that ham[ster]), was replaced by the initial part of the sentence that mentioned the monosyllabic word, up to and including the monosyllabic word itself (e.g. *ze dacht dat die ham [stukgesneden/steriel]*) or by the initial part of another recording of the carrier-word sentence. Thus, the experimental sentences all contained a spliced carrier word (e.g. *hamster*), but the first syllable of the carrier word

[1] As pointed out by an anonymous reviewer, theories of rhythm would predict that a stress clash between the successive stressed syllables would be avoided by applying the Silent Demibeat Addition or the Beat Addition rule, resulting in lengthening the first syllable or pausing between the two syllables (see Liberman & Prince, 1977; Selkirk, 1984).

originated from another token of the same carrier word or from a monosyllabic word. The different versions of cross-spliced sentences were therefore lexically identical; the critical difference between them was the acoustic realization of the ambiguous sequence. This manipulation ensured that any effect of the context from which the sequence originated would be independent of any effect due to subsequent disambiguating information.

We collected and analyzed the visual fixations to pictured objects that participants made as they listened to the cross-spliced sentences which mentioned one of the displayed objects (e.g. *ze dacht dat die hamster verdwenen was*, she thought that that hamster had disappeared). The participants' task was to click on and move the object referred to in the sentence with the computer mouse. Along with the target picture (e.g. the picture of a hamster), the picture associated with the monosyllabic word (e.g. *ham*), as well as two distractor pictures (e.g. *kraan* [tap] and *wasmachine* [washing machine], see Fig. 1) were presented. Because people usually fixate the object they intend to click on to guide the mouse movement, the fixations that participants perform as they hear the name of the target object reflect their current interpretation of the acoustic signal. This interpretation is taken to reflect the degree of lexical activation of potential word candidates. Allopenna, Magnuson, and Tanenhaus (1998) have shown that fixations to displayed pictures over time can be predicted from the lexical activation associated with the pictures' names as generated by a model like TRACE, given simple assumptions. The probability of fixating a pictured object has been shown to vary with the goodness of fit between the name of the picture and the spoken input computed at a very fine-grained level (Dahan, Magnuson, Tanenhaus, & Hogan, 2001), as well as with the lexical frequency associated with the picture's name (Dahan, Magnuson, & Tanenhaus, 2001). The eye-tracking paradigm therefore appears to offer a measure of lexical activation of potential candidates over time that could reflect subtle modulations as a function of the acoustic realization of an ambiguous sequence.

2. Experiment 1

Experiment 1 aimed to replicate and extend Davis et al. (2002) by testing whether the realization of an ambiguous sequence (e.g. /hɑm/, which could either be a monosyllabic word, *ham*, or the initial syllable of a carrier word, *hamster*) resulted in differential activation of the monosyllabic word and the carrier word. The visual target object was always the object corresponding to the carrier word; the competitor object was always the object representing the monosyllabic word. The acoustic realization of the carrier word was varied using cross-splicing: the first syllable of the carrier word was replaced by a recording of the monosyllabic word or by a different recording of the first syllable of the carrier word. In both cases, we predicted that as the target words unfolded over time, people would make more fixations to the competitor pictures than to the distractor pictures, thereby reflecting the strong match between the first syllable of the target word and the name associated with the competitor picture (i.e. the monosyllabic word). Of primary interest was whether participants' fixations to the competitor picture, as the ambiguous sequence was heard

and processed, differed across the splicing conditions. If the acoustic realization of the sequence conveyed disambiguating cues, we expected more fixations to the competitor picture when the sequence originated from a monosyllabic word than when it originated from a carrier word. This would suggest that the input provided more support for the monosyllabic interpretation of the sequence in the former case than in the latter.

Experiment 1 extended Davis et al. (2002) by varying the prosodic context in which the monosyllabic word was originally produced. In one version, the monosyllabic word was followed by a word stressed on its first syllable; in the other version, the monosyllabic word was followed by a word unstressed on its first syllable. Rakerd, Sennett, and Fowler (1987) showed that the duration of a monosyllabic word (e.g. *bike*) was longer when it was followed by an initially stressed word (e.g. *round*) than when it was followed by an initially unstressed word (e.g. *around*). We asked whether such a manipulation would affect the temporary lexical interpretation of the ambiguous sequence. The cross-spliced carrier words used in the eye-tracking experiment were constructed using the monosyllabic word produced in a stressed-syllable context (Experiment 1A) or in an unstressed-syllable context (Experiment 1B).

2.1. Method

2.1.1. Participants

Sixty native speakers of Dutch, students at the University of Nijmegen, participated in the experiment (30 in Experiment 1A, 30 in Experiment 1B).

2.1.2. Materials

Twenty-eight pairs of words were selected from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). Each word pair consisted of a carrier word and a monosyllabic word that phonemically matched the first (stressed) syllable of the carrier word. There were no semantic or morphological relationships between the monosyllabic and carrier words within each pair. All of these words were picturable nouns. Two additional picturable nouns were assigned to each word pair. These words were selected to be distractors presented along with the carrier and monosyllabic pictures in the eye-tracking experiment. The distractor words were phonologically dissimilar to the carrier word and the monosyllabic word. The 28 word pairs and their distractor words are listed in Appendix A. Pictures associated with the items were all black and white line drawings, selected from various picture databases (in particular, Cycowicz, Friedman, Rothstein, & Snodgrass, 1997; Snodgrass & Vanderwart, 1980).

Three sentences were constructed for every monosyllabic–carrier word pair: a sentence mentioning the carrier word and two sentences mentioning the monosyllabic word (see Table 1). The initial part of the sentence that preceded the carrier word or the monosyllabic word was identical for all three sentences and provided no semantic information indicating which of the carrier or the monosyllabic word was more likely to follow (e.g. *ze dacht dat die [hamster/ham]*, she thought that that [hamster/ham]). The monosyllabic word was always followed by a word that started with the same consonant or consonant cluster and the same vowel as the second syllable of the carrier word, with

Table 1
Example of a three-sentence set for one monosyllabic–carrier word pair used to produce the three versions of the cross-spliced sentence used in Experiment 1 (the underlined portion of each sentence was used to create the cross-spliced versions)

Carrier-word sentence	Zij dacht dat die ham*ster*$_a$ *verdwenen was*
	*Zij dacht dat die ham*ster$_b$ verdwenen was
	(She thought that that hamster had disappeared)
Monosyllabic-word sentence	
Stressed context	*Zij dacht dat die ham*$_c$ stukgesneden was
	(She thought that that ham had been sliced)
Unstressed context	*Zij dacht dat die ham*$_d$ steriel verpakt was
	(She thought that that ham had been wrapped under sterile conditions)
Cross-spliced sentences	
Carrier word	Zij dacht dat die ham$_b$ster$_a$ verdwenen was
Monosyllabic stressed-context	Zij dacht dat die ham$_c$ster$_a$ verdwenen was
Monosyllabic unstressed-context	Zij dacht dat die ham$_d$ster$_a$ verdwenen was
	(She thought that that hamster had disappeared)

the exception of the vowel /ʉ/, which was substituted for the reduced vowel /ə/ in four items in the unstressed-syllable context and in 18 items in the stressed-syllable context. (Note that these two vowels are very similar; Smits, Warner, McQueen, and Cutler (2003) have shown that they are perceptually highly confusable for Dutch listeners.) Depending on the condition, the word following the monosyllabic word was either stressed on its first syllable or not (e.g. *'stukgesneden* [sliced] or *ste'riel* [sterile]). In the former, the syllable always carried primary stress. In the latter, the syllable was unstressed in 23 out of the 28 items; for the remaining five items, the first syllable carried secondary stress. For contrast purposes, we will nevertheless refer to this condition as the unstressed-syllable condition. The sentences are listed in Appendix B.

All sentences were read aloud in a random order by a female speaker who did not know the purpose of the experiment, and recorded on DAT-tape in a sound-proof room. To induce a similar prosodic phrasing in all three sentences associated with each monosyllabic–carrier word pair, the speaker was instructed to produce the carrier word or the monosyllabic word as the focus of the sentence by accenting it. To this end, the monosyllabic word or the carrier word was marked on the script by the use of capitals. Each sentence was recorded successively at least four times. The sentences were then digitized, and edited and labeled using the Xwaves speech-editor. The specific recordings used to create the cross-spliced sentences were randomly selected from the available tokens, provided that they contained no disfluencies and could be spliced onto another sentence token without creating obvious acoustic artifacts. This mirrored Davis et al.'s (2002) stimulus selection procedure. There was no attempt to magnify or minimize the potential acoustic differences in the realization of the ambiguous sequence across conditions.

For each word pair, three cross-spliced sentences were created by splicing the initial portion of the carrier-word or monosyllabic-word sentences (up to and including the ambiguous sequence) with the same final portion of a different token of

the carrier-word sentence. These cross-spliced sentences were thus lexically identical to the carrier-word sentence, but differed in which sentence their initial portion originated from (i.e. the carrier-word sentence, the monosyllabic-word sentence in the stressed-context condition, or the monosyllabic-word sentence in the unstressed-context condition).[2]

Each experiment (i.e. Experiment 1A, comparing carrier-word and monosyllabic-word stressed-context conditions, and Experiment 1B, comparing carrier-word and monosyllabic-word unstressed-context conditions) contained 28 experimental trials. A trial consisted of the presentation of the pictures associated with one of the 28 word pairs and their distractors along with one of the three cross-spliced versions of the sentence. In addition, 42 filler trials were constructed. For each filler trial, a picturable word was selected to play the role of the target, along with three picturable distractor words (phonologically dissimilar to the target word). One important criterion for selecting the target words in the filler trials was the word's number of syllables. In all experimental trials, the target word was polysyllabic. To prevent participants from developing a possible bias toward target words being polysyllabic (which would have penalized monosyllabic-word interpretations of the ambiguous sequences), target words in filler trials were monosyllabic in 35 of the 42 trials, thus counterbalancing the number of monosyllabic and polysyllabic target words. Moreover, to prevent the possibility that participants might develop expectations that pictures with similar names were likely targets, 13 of the 42 filler trials had one distractor word embedded in the other distractor word (e.g. *trom* [drum] and *trompet* [trumpet]).

Pictures for the filler trials were selected from the same databases as were used for the experimental trials. In addition, sentences mentioning the filler target words were constructed. They were produced by the same speaker, and recorded at the same time as the experimental sentences. Cross-spliced filler sentences were created by concatenating two different recordings of a filler sentence. The initial part of one recording of each filler sentence, up to and including the monosyllabic target word or the first syllable of the polysyllabic target word, was spliced onto the final part of another recording of the same filler sentence, starting either at the word following the monosyllabic target word or at the second syllable of the polysyllabic target word.

[2] The splicing manipulation was done very carefully and did not create any obvious oddities that participants could easily detect while listening to the spliced versions of the sentences. To establish that spliced sentences were difficult to distinguish from their unspliced counterparts, we presented 18 participants (who did not participate in the eye-tracking experiment) with sentence pairs consisting of one of the three spliced versions of the carrier-word sentence and its original, unspliced counterpart (the token from which the last portion of the spliced sentence, constant across all three spliced versions, had been extracted). Participants were instructed to determine which one of those two lexically identical sentences had been artificially edited and manipulated. Participants heard all three possible pairings for each of the 28 experimental items; order of presentation was counterbalanced across participants. On average, the spliced sentence was accurately distinguished from its intact counterpart on 53.7% of the trials overall: 50.8% (ranging, across items, from 22% to 83%) when the initial portion of the spliced sentence originated from the carrier-word sentence, 56% (ranging from 33% to 83%) when it originated from the monosyllabic-word sentence in the stressed context, and 54.4% (ranging from 28% to 78%) when it originated from the monosyllabic-word sentence in the unstressed context. These results demonstrate that the spliced sentences were difficult to distinguish from intact sentences, and that the sentences did not have acoustic characteristics that rendered them readily detectable as manipulated speech.

2.1.3. Acoustic analyses

The duration of the sequences as well as the mean fundamental frequency (F0) of their vowels were measured to evaluate the extent to which the context in which sequences were produced affected their acoustic realization. On average, the duration of the ambiguous sequence was 245 ms when it originated from a carrier word, 265 ms when it corresponded to a monosyllabic word followed by a stressed syllable, and 259 ms when it corresponded to a monosyllabic word followed by an unstressed syllable. The differences in the ambiguous-sequence duration between the carrier- and monosyllabic-word conditions in the stressed-syllable context (stimuli used in Experiment 1A) ranged from −24 to 87 ms, with the monosyllabic-word sequence being longer than the carrier-word sequence for 25 of the 28 items. The differences in the ambiguous-sequence duration between the carrier and monosyllabic-word conditions in the unstressed-syllable context (stimuli used in Experiment 1B) ranged from −28 to 72 ms, with the monosyllabic-word sequence being longer than the carrier-word sequence for 22 of the 28 items. Consistent with what Davis et al. (2002) observed, this indicates that the sequence tended to be longer when corresponding to a monosyllabic word than to the first syllable of a carrier word, although the mean durational differences were substantially smaller here (20 and 15 ms) than in the Davis et al. (2002) study (48 ms). Measures of the mean F0 value of the vowels in each sequence revealed a negligible effect of the context in which the sequence was produced (264 Hz in the carrier-word condition, 267 Hz in the monosyllabic-stressed context condition, and 265 Hz in the monosyllabic-unstressed context condition).

2.1.4. Procedure and design

Prior to the eye-tracking experiment, participants were familiarized with the pictures to ensure that they identified and labeled them as intended. Each picture appeared on a computer screen in the same format as that used in the eye-tracking experiment, along with its printed name. Participants were instructed to familiarize themselves with each picture and to press a response button to proceed to the next picture. After this part of the experiment, the eye-tracking system was set up.

Participants were seated at a comfortable distance from the computer screen. One centimeter on the visual display corresponded to approximately 1° of visual arc. The eye-tracking system was mounted and calibrated. Eye movements were monitored with an SMI Eyelink eye-tracking system, sampling at 250 Hz. Spoken sentences were presented to the participants through headphones. The structure of a trial was as follows. First, a central fixation point appeared on the screen for 500 ms, followed by a blank screen for 600 ms. Then, a 5 × 5 grid with four pictures and four geometrical shapes appeared on the screen (see Fig. 1) as the auditory presentation of a sentence was initiated. Prior to the experiment, participants were instructed to move the object mentioned in the spoken sentence above or below the geometrical shape adjacent to it, using the computer mouse. The positions of the pictures were randomized across four fixed positions of the grid while the geometrical shapes appeared in fixed positions on every trial. Participants' fixations for the entire trial were completely unconstrained and participants were under no time pressure to perform the action. The position of the mouse cursor on the computer screen while the mouse button was pushed (i.e. while the object was picked up and moved) was sampled and recorded, along with the eye-movement data. The software controlling

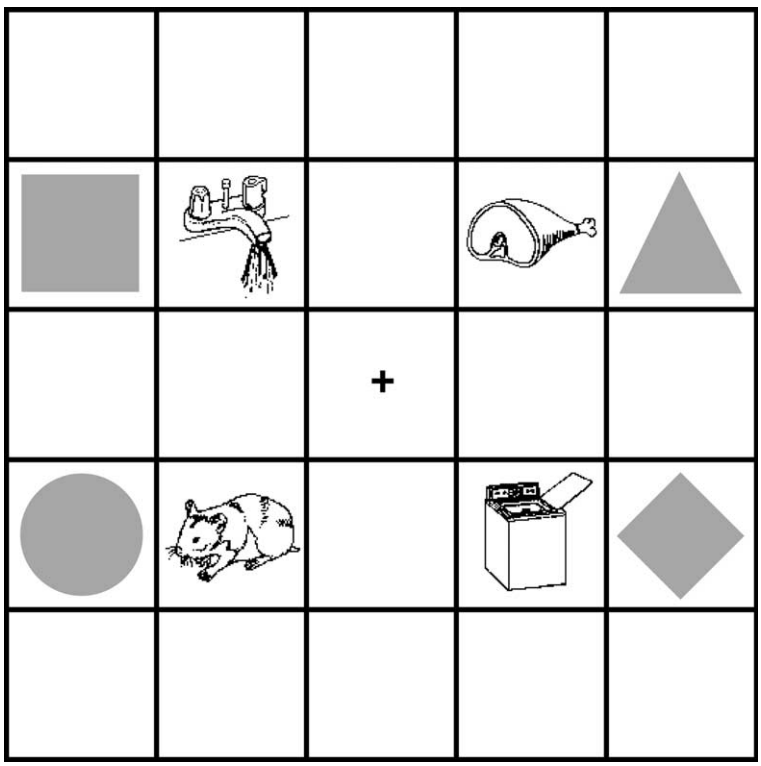

Fig. 1. Example of a visual display. The geometrical shapes were green.

stimulus presentation (pictures and spoken sentences) interacted with the eye-tracker output so that the timing of critical events in the course of a trial (such as the onsets of the spoken stimuli and mouse movements) was added to the stream of continuously sampled eye-position data. Once the picture had been moved, the experimenter pressed a button to initiate the next trial. Every five trials, a central fixation point appeared on the screen, allowing for some automatic drift correction in the calibration.

Within each experiment (Experiment 1A or 1B), two lists were created by varying which of the two versions of the spliced sentences (monosyllabic word or carrier word) was presented for each of the 28 experimental items. Within each list, 14 experimental items were assigned to each condition. For each list, eight random orders were created, with the constraint that five of the filler trials were presented at the beginning of the experiment to familiarize participants with the task and procedure. Participants were randomly assigned to each list, with an approximately equal number of participants assigned to each random order.

2.1.5. Coding procedure

The data from each participant's right eye were analyzed and coded in terms of fixations, saccades, and blinks, using the algorithm provided in the Eyelink software. (For a few participants, data for the left eye were used because of calibration problems with

the right eye.) Onsets and offsets of saccades are automatically determined using the default thresholds for motion (0.2 degrees), velocity (30 degrees/s), and acceleration (8000 degrees/s^2). Fixation durations correspond to the time intervals between two successive saccades and fixation positions were determined by averaging the *x* and *y* coordinates of the eye positions recorded during the fixation. The timing of the fixations was established relative to the onset of the target word in the spoken utterance. Graphical analysis software performed the mapping between the position of fixations, the mouse movements, and the pictures present on each trial, and displayed them simultaneously. Each fixation was represented by a dot associated with a number which denoted the order in which the fixation had occurred; the onset and duration of each fixation were available for each fixation dot.

For each experimental trial, fixations were coded from the onset of the target word until participants had clicked on the target picture with the mouse, which was taken to reflect the participants' confident identification of the target word. In most cases, participants were fixating the target picture when clicking on it. In the rare cases where participants clicked on the target picture long after the offset of the target word and/or when not simultaneously looking at the target picture, an earlier long fixation to the target picture was taken as indicating recognition of the target word. Fixations were coded as directed to the target picture (always the picture associated with the carrier word), to the competitor picture (always the picture associated with the monosyllabic word), to one of the two distractor pictures, or to anywhere else on the screen. Fixations that fell within the cell of the grid in which a picture was presented were coded as fixations to that picture.

2.2. Results

The goal of Experiment 1 was to examine whether the degree to which the competitor picture associated with a monosyllabic word (e.g. the picture of a ham) was considered, as the target word (e.g. hamster) was heard and processed, depended on the word from which the first syllable of the cross-spliced target word originated. We compared conditions in which the first syllable of the target word came from another token of the carrier word and from a monosyllabic word followed by a stressed syllable (Experiment 1A), or from the same token of the carrier word and from a monosyllabic word followed by an unstressed syllable (Experiment 1B).

2.2.1. Experiment 1A

On a few trials, participants erroneously moved the competitor picture instead of the target picture without correcting their choice (13 out of 840 trials, 1.5% of the data). These trials were excluded from the analyses. The proportion of fixations to each picture or location (i.e. target picture, competitor picture, distractor pictures, or elsewhere) over time (in 10 ms time intervals) for each condition and each participant was calculated by adding the number of trials in which a picture type was fixated during a specific time interval and dividing it by the total number of trials where a fixation to any picture or location was observed during this time interval (thus excluding in this count the trials where a blink or a saccade occurred during that time interval).

Fig. 2 presents the average proportion of fixations, across participants, to each type of picture (target, competitor, or averaged distractors) from 0 to 1000 ms after the onset of

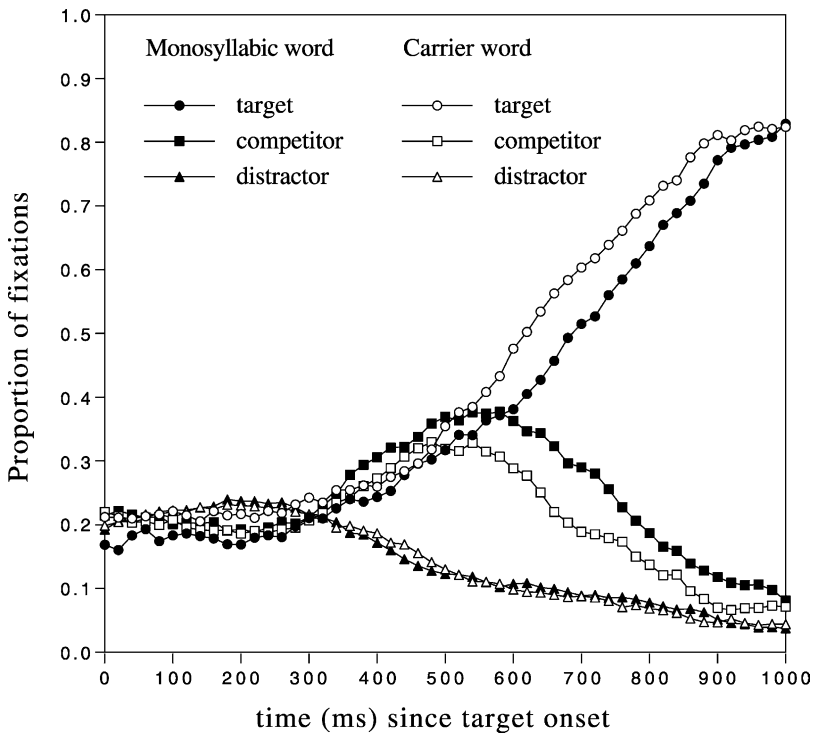

Fig. 2. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 1A (carrier-word vs. monosyllabic-word stressed-context condition).

the target word. As is apparent from the graph, the proportions of fixations to any picture on the display were equivalent at the target-word onset, demonstrating no fixation bias before any relevant information about the target picture was heard. Around 300 ms, fixation proportions to the target picture began to rise in both conditions and steadily increased until they reached about 0.85 by 1000 ms. Conversely, fixation proportions to the distractor pictures decreased steadily from 300 to 1000 ms. This indicates that the mapping of the signal onto lexical representations is reflected by fixations from 300 ms on. Given an estimate of 200 ms for programming a saccade (Hallett, 1986), fixations occurring at 300 ms were programmed after hearing about 100 ms of the target word. Fixation proportions to the competitor picture began to increase at 300 ms in both conditions and in parallel to the fixations to the target picture. Importantly, the fixation proportion to the competitor picture increased faster, reached a higher peak, and decreased more slowly in the monosyllabic-word condition than in the carrier-word condition. This demonstrates that the realization of the ambiguous sequence (as captured by the word it originated from) modulated the degree to which the competitor picture was considered. Fixation proportions to the target picture across conditions showed the mirror image of this effect. The fixation proportion to the target picture rose faster in the carrier-word condition than in the monosyllabic-word condition.

The difference between conditions was statistically tested by computing the average fixation proportion to the competitor picture over a time window extending from 300 to 900 ms. Analyses of variance (ANOVAs) were performed on these fixation proportions with participants (F_1) and with items (F_2) as the repeated measures. The 300–900 ms time window corresponded to the interval over which fixation proportions to the competitor picture were higher than fixation proportions to the distractor pictures. Over this time interval, the average proportion of fixations to the competitor picture was 28% in the monosyllabic-word condition and 23% in the carrier-word condition. A one-way ANOVA (monosyllabic condition vs. carrier condition) indicated that this difference was reliable ($F_1(1,29) = 11.6$, $P < 0.005$; $F_2(1,27) = 5.5$, $P < 0.05$).

A notable aspect of the data concerns the time interval over which the difference in competitor fixations between the monosyllabic-word and carrier-word conditions was largest. As is apparent in Fig. 2, this difference between conditions was modest early on and became large later in time. Considering that the target words in the monosyllabic-word and carrier-word conditions differed in their ambiguous sequence only, one may have expected to observe a larger effect of the realization of the ambiguous sequence between 300 and 550 ms, that is, during the time window over which this sequence, of about 250 ms, was heard and processed. However, such an expectation is based on the assumption that the acoustic realization of the ambiguous sequence would contain specific acoustic cues biasing its interpretation. The observed pattern suggests that these signals occurred late in the sequence, and/or that the interpretation of the ambiguous sequence was biased by information accumulating over time, rather than by discrete cues favoring one interpretation or the other.

In order to evaluate whether the size of the effect was reliably stronger after rather than while the ambiguous sequence was processed, we conducted a two-way (Condition × Time Window [300–550 ms vs. 550–900 ms]) ANOVA. The difference in competitor fixation proportion across the monosyllabic- and carrier-word conditions was small between 300 and 550 ms (31% in the monosyllabic-word vs. 28% in the carrier-word condition) but large between 550 and 900 ms (26% vs. 19%). There was a main effect of Condition ($F_1(1,29) = 9.6$, $P < 0.005$; $F_2(1,27) = 4.9$, $P < 0.05$), and a main effect of Window ($F_1(1,29) = 23.2$, $P < 0.001$; $F_2(1,27) = 10.1$, $P < 0.005$), but the interaction did not reach significance ($F_1(1,29) = 1.9$, $P > 0.10$; $F_2(1,27) = 3.1$, $P > 0.05$). Thus, this analysis does not provide compelling evidence that the effect of the cross-splicing manipulation changed over time.

An additional aspect of the data as shown in Fig. 2 is noteworthy: the time interval over which the fixation proportion to the competitor was higher than that to the distractors. The interval extended for about 600 ms (i.e. from 300 ms up to 900 ms), even in the carrier-word condition. As is apparent in Fig. 2, fixations to the competitor picture began to increase around 300 ms, and began to decrease between 550 and 600 ms after target onset, thus between 250 and 300 ms after the point at which fixations start to reflect the uptake of the critical acoustic information. The duration of the ambiguous sequence was approximately 250 ms (245 ms in the carrier-word condition and 265 ms in the monosyllabic-word condition). Thus, the drop in competitor fixations at this point reflects the fact that, after the ambiguous sequence, the signal continued to provide support for a carrier-word interpretation (e.g. the sequence /stər/ being consistent with the “hamster” interpretation),

thus accumulating more evidence in favor of the target picture, to the detriment of the competitor picture. However, competitor fixations remained quite high for an extended amount of time after the point where they started to drop, that is, from 550–600 to 900 ms. This time interval, over which the competitor fixations decreased before they merged with the distractor fixations, appears to be larger than those found in past eye-tracking studies examining the activation of cohort-like competitors, such as the activation of *beetle* when the target word *beaker* is heard (Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Assuming that the time window over which competitor fixations remain higher than distractor fixations reflects the time course of competitor activation, the activation of the competitor (which corresponds to the monosyllabic word embedded in the target word) remained high for a substantial amount of time after it started to decrease. We will return to this point in Section 5.

2.2.2. Experiment 1B

Experiment 1B was identical to Experiment 1A in all aspects except for the ambiguous sequences used in the monosyllabic-word condition. Here, these sequences had been produced as monosyllabic words followed by an unstressed syllable.

On a few trials, participants erroneously moved the competitor picture instead of the target picture without correcting their choice (15 out of 840 trials, 1.8% of the data). These trials were excluded from the analyses. Fig. 3 presents the fixation proportions to the target picture, the competitor picture, and to the averaged distractor pictures, from 0 to 1000 ms after the onset of the target word. At the onset of the target word, fixation proportions to various pictures did not differ. Around 300 ms after target onset, fixation proportions to the target and competitor pictures began to increase, while those to the distractor pictures began to decrease. Fixation proportions to the competitor picture remained higher than those to the distractor pictures until around 900 ms, where they merged again. This pattern is consistent with what was found in Experiment 1A. However, the difference in competitor and target fixations between the carrier-word and the monosyllabic-word conditions, although in the same direction, was noticeably smaller than that found in Experiment 1A.

The fixation proportion to the competitor picture, averaged over the 300–900 ms time window, was 27% in the monosyllabic-word condition and 24% in the carrier-word condition. A one-way ANOVA (monosyllabic condition vs. carrier condition) on the average fixation proportions revealed that this difference was significant by participants but not by items ($F_1(1,29) = 5.9$, $P < 0.05$; $F_2(1,27) = 1.5$, $P > 0.10$), suggesting large variability across items. A two-way (Condition × Time Window [300–550 ms vs. 550–900 ms]) ANOVA revealed a significant effect of Window ($F_1(1,29) = 65.7$, $P < 0.001$; $F_2(1,27) = 19.1$, $P < 0.001$), an effect of Condition significant only by participants ($F_1(1,29) = 5.2, P < 0.05; F_2(1,27) = 1.4, P > 0.10$), and no interaction ($F_1$ and $F_2 < 1$).

In order to compare the pattern of results from Experiments 1A and 1B, a two-way (Condition × Experiment) ANOVA was conducted over the 300–900 ms time window. Experiment was treated as a between-subjects factor in the F_1 analysis and as a within-items factor in the F_2 analysis. There was a main effect of Condition ($F_1(1,58) = 17.4$, $P < 0.001$; $F_2(1,27) = 4.8$, $P < 0.05$), no main effect of Experiment, and no interaction between the two factors. Thus, the stress status of the syllable following the monosyllabic word does not appear to have a systematic impact on lexical disambiguation. However,

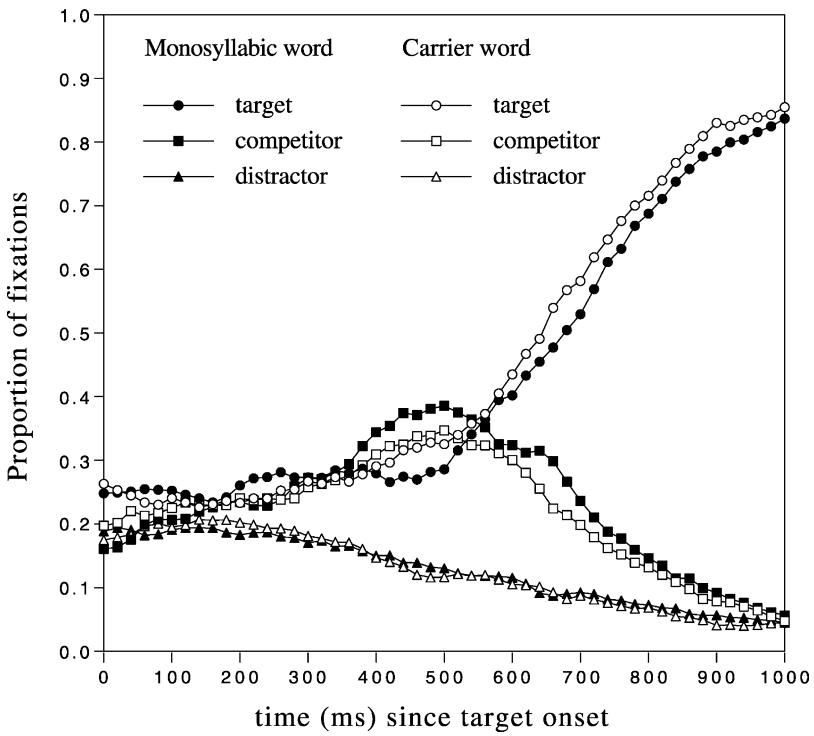

Fig. 3. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 1B (carrier-word vs. monosyllabic-word unstressed-context condition).

the inter-item variability across items observed in Experiment 1B but not in Experiment 1A (with the same sampling procedure and statistical power in both experiments) suggests that embedding disambiguation is determined by another factor than the lexical origin of the ambiguous sequence.

2.3. Discussion

Experiment 1 examined whether the acoustic realizations of a monosyllabic word and the first syllable of its carrier word differ in a way that affects lexical interpretation. Using cross-splicing, we presented participants with lexically and phonemically identical sentences containing a carrier word (e.g. *hamster*). However, the first syllable of that word, that is, the ambiguous sequence, originated from another recording of the carrier word or from the recording of a monosyllabic word. This manipulation was realized with the monosyllabic word originally followed by a stressed syllable (Experiment 1A) and by an unstressed syllable (Experiment 1B).

Experiment 1A showed that participants fixated the competitor picture representing the monosyllabic-word interpretation of the ambiguous sequence more when that ambiguous sequence originated from the recording of a monosyllabic word than when it

originated from the recording of a carrier word. This demonstrates that a phonemically identical sequence can contain cues that modulate its interpretation. This is an important result because it confirms that listeners' uptake of information from the acoustic signal cannot be captured by a purely phonemic description of the sequence. This finding is consistent with what Davis et al. (2002) reported, using a different task and different materials.

Experiment 1B showed a similar pattern of results, but the bias in interpreting the ambiguous sequence as a monosyllabic word when it originated from a monosyllabic word was numerically reduced and not significant by items. This is reflected in the visual inspection of Figs. 2 and 3: the difference in competitor fixations between the monosyllabic- and carrier-word conditions was smaller in Experiment 1B than in Experiment 1A. The non-significant interaction between Experiment and Condition, however, suggests that the stress status of the following syllable is a prosodic factor that does not reliably influence the lexical interpretation of the ambiguous sequence. Nevertheless, the failure to find a robust effect of the splicing manipulation in Experiment 1B, with the same statistical power as Experiment 1A and closely matched stimuli, is important because it indicates that the lexical disambiguation of an embedded sequence may not be as systematic a phenomenon as Davis et al. (2002) concluded. It also challenges the suggestion that the acoustic cues that contribute to this disambiguation are lexically determined (i.e. are stored lexically in the speech production system). This is because such an account does not predict variability – other than noise – in the production of disambiguating cues.

One way of accounting for this variability, as we suggested in Section 1, is to assume that the lexical disambiguation of an ambiguous sequence is influenced by the presence and/or strength of a prosodic boundary following a monosyllabic word, rather than by the mere production of a monosyllabic or longer word. The realization of a monosyllabic word may differ from that of the first syllable of a carrier word because a major prosodic-constituent boundary is likely to follow the former, but not the latter. Recall that the sequence was longer, on average, when produced as a monosyllabic word than as a carrier word, and slightly longer when the monosyllabic word was followed by a stressed syllable than by an unstressed syllable. If sequence duration is taken as an index of the presence and/or strength of a prosodic boundary (e.g. Beckman & Edwards, 1990; Turk & Shattuck-Hufnagel, 2000), the phonetic correlates of a prosodic boundary were often produced when the sequence corresponded to a monosyllabic word, but not when the sequence corresponded to the first syllable of a longer word. Likewise, a prosodic boundary may have been more often or more strongly marked in the utterances selected in the monosyllabic-word stressed-context condition than in those selected in the monosyllabic-word unstressed-context condition. This hypothesis also assumes that the acoustic correlates of a prosodic boundary, such as segmental lengthening,[3] are used

[3] The term "segmental lengthening" implies a reference duration, and the computation of such reference almost certainly involves the preceding prosodic context in which the lengthened sequence occurs. For example, durational lengthening of a sequence could be assessed after establishing that its segments are longer than what would be expected given, for instance, the speaker's speech rate. However, because we lack a model of how such a reference duration is computed, we will use the absolute duration of the sequence as an estimate of its relative value.

probabilistically by listeners. The larger the boundary, as characterized by its acoustic correlates, the larger the bias to interpret the sequence as corresponding to an embedded, monosyllabic word.[4]

In order to evaluate the prosodic-boundary hypothesis, we computed the correlation over the 28 items between the difference in duration between the monosyllabic-word and carrier-word sequences and the difference in competitor fixations between the monosyllabic-word and carrier-word conditions over the 300–900 ms time window, thus factoring out item- and picture-dependent variability. A very strong relationship between these two measures was observed (for Experiment 1A: $r(26) = 0.61$, $P < 0.001$; for Experiment 1B: $r(26) = 0.54$, $P < 0.005$; for both experiments: $r(54) = 0.59$, $P < 0.001$). These correlations suggest that the degree to which the competitor picture is considered is related to the duration of the ambiguous sequence, which, we argue, reflects the strength of a prosodic boundary. The longer the sequence, the more it is interpreted as a monosyllabic word. This is consistent with our claim: a lexical-interpretation bias would result from the presence of acoustic characteristics associated with a prosodic boundary, such as durational lengthening. Interestingly, Davis et al. (2002) reported a significant correlation between the magnitude of durational and F0 differences between monosyllabic- and carrier-word stimuli (from naive and non-naive speakers) and listeners' ability at predicting which word the ambiguous sequence originated from. They suggested that this relationship reflects the *additional* contribution to disambiguation of prosodic-boundary cues after the monosyllabic words, produced by the naive speakers but not by the non-naive speaker. In our view, there is only one factor responsible for lexical-embedding disambiguation, namely, the production of prosodic boundaries, which manifests itself in a variable and gradient manner. This naturally explains the effect of the origin of the sequence (from a monosyllabic or carrier word) on its interpretation.

Before pursuing our enterprise of validating the prosodic-boundary hypothesis, an alternative account of our results needs to be considered. This account hinges on the interdependency between duration and processing time. Zwitserlood and Schriefers (1995) demonstrated that the degree of activation of a word increases as the length of the portion of the signal consistent with it increases, but also as more time for processing a short portion of the signal is allowed. This suggests that activation accrues over time, even in the absence of additional bottom-up support. A long ambiguous sequence would thus allow the activation of all candidates that are consistent with it to accrue more than a shorter sequence would, until the signal disambiguates between the candidates. This predicts

[4] An alternative explanation for the difference in lexical disambiguation between Experiments 1A and 1B bears on the influence of coarticulatory information from the following context on the sequence's realization. While the consonant or consonant cluster following the sequence was exactly matched across all three conditions (e.g. the sequence "ham" was followed by "st" in the carrier-word, the monosyllabic-stressed context condition and the monosyllabic-unstressed context condition), the following vowel was not always identical. The reduced vowel /ə/ in the carrier-word condition was substituted by the full vowel /ʉ/ in 18 out of the 28 items in the monosyllabic-stressed context condition, but only in four items in the monosyllabic-unstressed context condition. Coarticulation of these context vowels with the sequence vowels might have differentially affected the realization of the sequence vowels, thus providing listeners with non-durational cues to lexical interpretation. This alternative explanation can be rejected on the basis of the results of Experiments 2 and 3, where the duration of the sequence, rather than the context in which it was originally produced, biased its lexical interpretation.

higher activation levels for *all* words consistent with the input when the duration of the input increases. This could account for higher fixation proportions to the competitor picture for long ambiguous sequences than for short ambiguous sequences.

The fact that lower fixation proportions to the target picture were observed when the ambiguous sequence was longer than when it was shorter seems at first incompatible with an explanation of the present results in terms of an increase of lexical activation with increased processing time. This is because more processing time should equally benefit the activation of all consistent words. However, active candidates inhibit each other proportionally to their own activation, and word activation varies with the word's lexical frequency. As the activation of frequent words increases with processing time, the activation of less frequent competitors decreases. In this experiment, and in the Dutch language in general, short words tend to be more frequent than their carrier words. The more active short words are, the more they can inhibit their long, carrier competitors, resulting in lower fixation proportions to the target (carrier) pictures as fixation proportions to the competitor (monosyllabic) pictures increase. Averaged across items, our results are compatible with this alternative account. However, a number of analyses conducted on Experiment 1A's results provide no support for this account. In particular, when looking at the few items for which the frequency of the target (carrier) word (on the basis of the CELEX database) could reliably be assessed as being higher than that of the competitor (monosyllabic) word (namely, *kei-kijker*, *lei-leiding*, *schil-schilder*, *sla-slager*, and *pin-pinda*), fixation proportions to the target over time were lower when the sequence durations were longer than when the sequences were shorter. This is the reverse of what the account based on an increase of lexical activation with increased processing time, in interaction with frequency, would predict. Furthermore, there were weak and non-significant correlations between the difference in frequency between the target (carrier) word and the competitor (monosyllabic) word and the size of the effect (i.e. the difference between carrier- and monosyllabic-word conditions) on target fixations in the 300–900 ms time interval ($r(26) = -0.02$), and on competitor fixations in this interval ($r(26) = 0.09$). There is thus no supporting evidence for an account of our results in which an increase of competitor activation would result from an increase in processing time for longer sequences.

In order to further examine how systematically the production of monosyllabic words or longer words provides disambiguating information, we replicated Experiment 1A with different spoken stimuli. We evaluated the lexical interpretation of an ambiguous sequence as a function of the context in which it originally occurred (i.e. in a carrier word or as a monosyllabic word followed by a stressed syllable). However, in contrast with Experiment 1A, we specifically selected the tokens used to create cross-spliced carrier words such that, for each item, the difference in the ambiguous-sequence duration between the carrier-word and monosyllabic-word conditions was minimized (Experiment 2) or opposite to Experiment 1A's pattern (Experiment 3). These manipulations directly tested the claim that the duration of an ambiguous sequence, more than the word it originates from, governs its lexical interpretation. Such a role of sequence duration would be consistent with the hypothesis that the disambiguation of lexical embedding mostly depends on the presence of acoustic cues such as segmental lengthening that mark the presence of a prosodic boundary.

3. Experiment 2

Experiment 2 evaluated the lexical interpretation of an ambiguous sequence that originated from a carrier word or a monosyllabic word when the sequence's duration was held constant between these conditions. Under the assumptions that (a) the durational lengthening of the segments of a sequence can be taken as an estimate of the presence and/or strength of a prosodic boundary following the sequence, and (b) the presence of a prosodic boundary results in a bias in favor of lexical candidates whose word boundaries are aligned with the hypothesized prosodic boundary, we predicted that eliminating the sequence-duration difference associated with the context in which the sequence was produced (monosyllabic or carrier word) would result in reducing or even eliminating the effect of this context on the lexical interpretation of the sequence.

3.1. Method

3.1.1. Participants

Thirty native speakers of Dutch, all students at the University of Nijmegen, took part in the experiment. None of them had participated in Experiments 1A or 1B.

3.1.2. Materials and procedure

Our stimuli were selected from the same source as the stimuli used in Experiment 1A. Over all the tokens available from our original recording, the duration of the ambiguous sequence was 248 ms ($N = 120$, SD = 42 ms) when it originated from a carrier word and 253 ms ($N = 142$, SD = 40 ms) when it corresponded to a monosyllabic word followed by a stressed syllable. As these numbers make clear, the two distributions of sequence duration overlapped to a great extent. Specific tokens of the carrier- and monosyllabic-word sentences were selected from the original recording such that the sequence-duration difference between the two sentence types, for each of the 28 items, was as small as possible. The average duration of the sequence was 248 ms (SD = 42 ms) in the carrier-word condition and 250 ms (SD = 40 ms) in the monosyllabic-word condition. The difference in the sequence duration across conditions was thus 2 ms on average, ranging from −4 to 32 ms. For 22 of the 28 items, the difference was less than 5 ms. The averaged values in both conditions were very similar to the averaged sequence duration in the carrier-word condition of Experiment 1A (245 ms). Measures of the mean F0 value on the sequences' vowel showed a negligible difference between the conditions (265 and 264 Hz in the carrier-word and the monosyllabic-word conditions, respectively).

Cross-spliced sentences were created using the same procedure as in Experiment 1. Design, procedure, and coding were the same as in Experiment 1.

3.2. Results and discussion

Fifteen trials were excluded from the analysis, either because participants erroneously moved the competitor picture without correcting their choice (12 out of 840 trials, 1.4% of the data) or because participants did not fixate the target picture before moving it (three trials, 0.4% of the data). Fig. 4 presents the proportion of fixations over time to the target

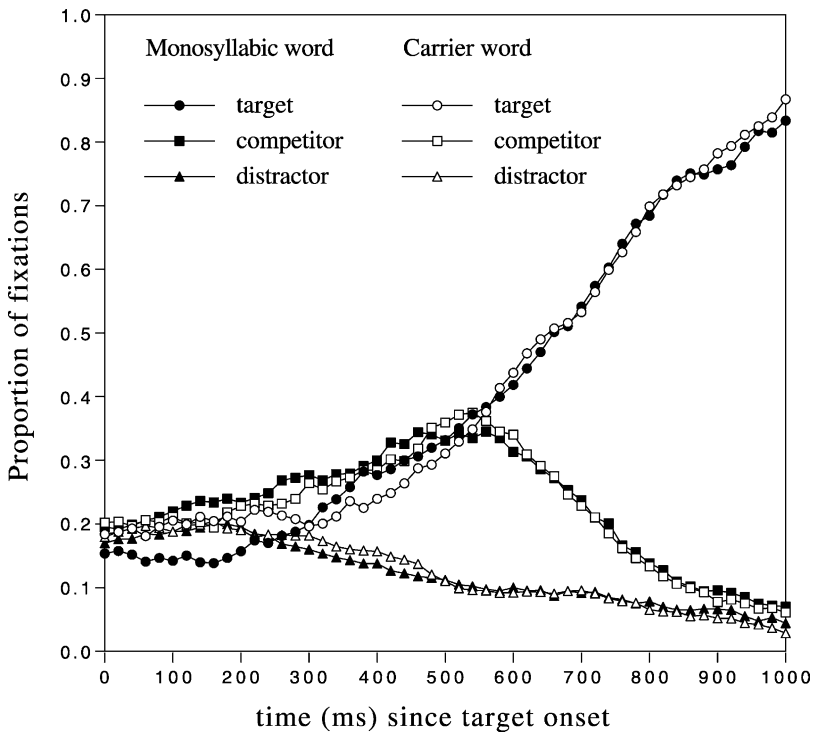

Fig. 4. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 2.

picture, the competitor picture, and to the averaged distractor pictures. As is immediately apparent from the graph, the fixation proportions to the target and competitor did not differ across conditions. In both conditions, fixation proportions to target and competitor began to rise while fixation proportions to the distractors began to decrease around 200 ms after the target-word onset, thus slightly earlier than in Experiment 1. Fixations to the competitor remained higher than to the distractors until around 900 ms.

The average fixation proportions to the competitor picture, computed over a 300–900 ms time window, confirmed this visual impression. The proportion of fixations to the competitor picture was 25% in the carrier-word condition and 25% in the monosyllabic-word condition. A one-way (carrier vs. monosyllabic) ANOVA confirmed the absence of an effect of Condition ($F_1 < 1$; $F_2 < 1$). A two-way (Condition × Experiment) ANOVA on fixation proportions to the competitor picture over the 300–900 ms interval was conducted in order to compare the results of Experiment 1A and Experiment 2. Experiment was treated as a between-subjects factor in the F_1 analysis and as a within-items factor in the F_2 analysis. The analysis revealed a significant effect of Condition, although this effect was marginal by items ($F_1(1,58) = 4.2$, $P < 0.05$; $F_2(1,27) = 3.4$, $P = 0.08$), no main effect of Experiment, and, importantly, a significant interaction between Condition and Experiment ($F_1(1,58) = 4.0$, $P < 0.05$; $F_2(1,27) = 4.2$, $P = 0.05$).

In Experiment 2, participants were thus equally likely to fixate the competitor picture whether the ambiguous sequence was originally produced as a monosyllabic word or as the first syllable of a carrier word. This is in sharp contrast with Experiment 1A's results, even though the conditions were defined and operationalized in identical terms. The only difference between these two experiments was whether the tokens used to construct cross-spliced sentences were randomly chosen or specifically selected in terms of the duration of the ambiguous sequence. When the duration of the sequence was matched between the monosyllabic-word and carrier-word conditions and equally short, there was no influence of the origin of the ambiguous sequence on its lexical interpretation.

This result shows that the production of monosyllabic or longer words does not always disambiguate between the two lexical interpretations. This finding, and the evidence from our recording that the sequence-duration distributions from monosyllabic and carrier words overlap to a large extent, call into question the possibility that the production of disambiguating cues to onset embedding is lexically determined. By contrast, the present results are in agreement with our claim that lexical interpretation is modulated by the presence of acoustic correlates to prosodic boundaries, such as sequence lengthening. If an ambiguous sequence is long, as in the monosyllabic-word condition from Experiment 1A, lexical candidates that require a word boundary aligned with the phonetically marked prosodic boundary are favored. When the sequence is short, as in both conditions in Experiment 2, no bias in lexical interpretation is observed.

4. Experiment 3

Experiment 3 aimed to provide a stronger test of the hypothesis that the presence of prosodic boundaries, as acoustically marked by segmental lengthening, favors lexical candidates whose edges are aligned with such boundaries. We selected sequence tokens such that the tokens produced as a monosyllabic word (followed by a stressed syllable) were shorter than the tokens produced as the first syllable of a carrier word. The sequence-duration pattern in Experiment 3 was thus reversed from the pattern present in Experiment 1A's stimuli and from the overall pattern in our recording. If the duration of the sequence, as an index of a prosodic boundary, determines the degree to which a monosyllabic-word interpretation is considered, we predicted that we would observe more fixations to the competitor picture (associated with the monosyllabic-word interpretation) when the ambiguous sequence was long but originated from a carrier word than when the sequence was short but corresponded to a monosyllabic word.

4.1. Method

4.1.1. Participants

Thirty native speakers of Dutch, all students at the University of Nijmegen, took part in the experiment. None of the students had participated in any of the previous experiments.

4.1.2. Materials and procedure

New cross-spliced stimuli were created by selecting from the original recording tokens for which the ambiguous sequence had the longest duration when it had been produced as part of a carrier word and tokens for which the sequence had the shortest duration when it had been produced as a monosyllabic word followed by a stressed syllable. As a result, the carrier-word sequence was longer than the monosyllabic-word sequence for 21 out of the 28 items (267 ms [SD = 42 ms] vs. 236 ms [SD = 42 ms], with duration differences between the two conditions ranging from 8 to 73 ms). For the remaining seven items, the sequence was always longer (or of an equal duration) when produced as a monosyllabic word than when produced as part of a carrier word. These seven items were included in the experiment, but excluded from all analyses. There was a negligible difference in the mean F0 on the sequences' vowels between the monosyllabic-word condition (261 Hz) and the carrier-word condition (266 Hz). Design, procedure, and coding were identical to Experiments 1 and 2.

4.2. Results and discussion

On a few trials, participants erroneously moved the competitor picture rather than the target picture (three out of 630 trials, 0.5% of the data). These trials were excluded from the analyses. Fig. 5 presents the proportion of fixations to the target picture, to the competitor

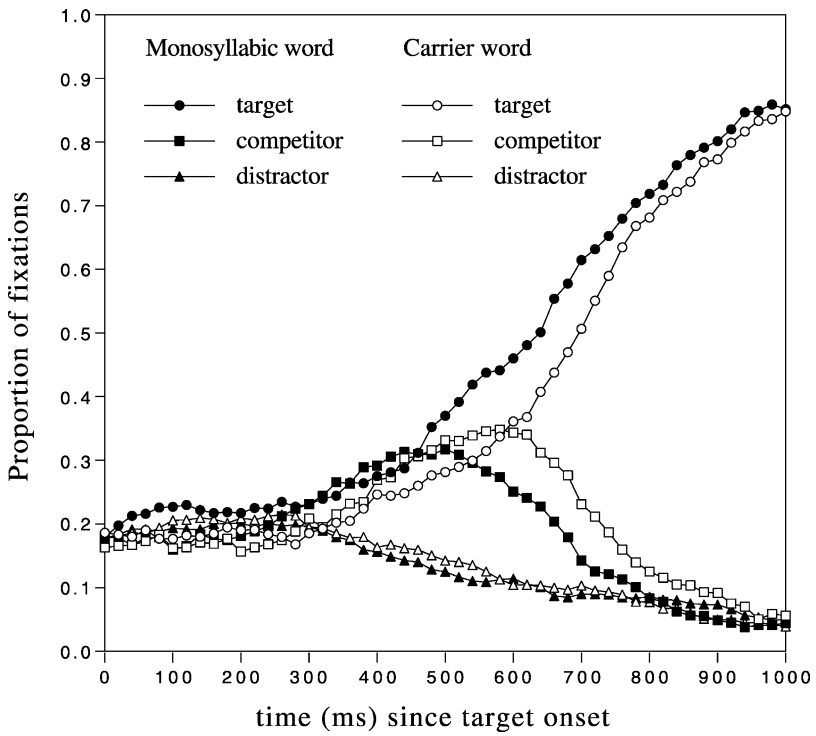

Fig. 5. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 3.

picture, and to the averaged distractor pictures over time, from 0 to 1000 ms after the onset of the target word. As in the previous experiments, at around 300 ms, target and competitor fixation proportions began to rise and distractor fixation proportions began to decrease. There was a major effect of conditions such that, around 550 ms after target-word onset, participants tended to fixate the competitor picture more when the ambiguous sequence originated from a carrier word but was of a long duration than when it originated from a monosyllabic word but was of a short duration.

Over the 300–900 ms time window, the average proportion of fixations to the competitor picture was 21% in the monosyllabic-word condition and 24% in the carrier-word condition. A one-way ANOVA showed that this effect was statistically not significant ($F_1(1,29) = 2.2$, $P > 0.10$; $F_2(1,20) = 1.5$, $P > 0.10$). A two-way (Condition × Time Window [300–550 ms vs. 550–900 ms]) ANOVA revealed no main effect of Condition ($F_1(1,29) = 1.2$, $P > 0.10$; $F_2(1,20) < 1$), a main effect of Window ($F_1(1,29) = 22.8$, $P < 0.001$; $F_2(1,20) = 16.5$, $P < 0.005$) and, crucially, a significant interaction ($F_1(1,29) = 4.6$, $P < 0.05$; $F_2(1,20) = 6.7$, $P < 0.05$). The difference in competitor fixations was small and not significant over the 300–550 ms time window (29% in the monosyllabic-word condition and 27% in the carrier-word condition; $F_1 < 1$; $F_2 < 1$), but large and significant between 550 and 900 ms (15% vs. 22%; $F_1(1,29) = 8.5$, $P < 0.01$; $F_2(1,20) = 7.4$, $P < 0.05$). There was also a significant correlation between the difference in duration between the monosyllabic-word and carrier-word conditions and the difference in the competitor fixation proportion over the 550–900 ms interval between these two conditions ($r(19) = 0.54$, $P < 0.01$; this correlation was also significant for the 300–900 ms time interval, $r(19) = 0.72$, $P < 0.001$).

A two-way (Condition × Experiment) ANOVA on the fixation proportions to the competitor picture over the 550–900 ms interval was conducted, comparing the results of Experiments 1A and Experiment 3, after excluding from the Experiment 1A data the seven items that were excluded from the Experiment 3 analyses. Experiment was treated as a between-subjects factor in the F_1 analysis and as a within-items factor in the F_2 analysis. The analysis revealed a significant effect of Experiment ($F_1(1,58) = 8.8$, $P < 0.005$; $F_2(1,20) = 11.7$, $P < 0.005$), a non-significant effect of Condition, and a significant interaction ($F_1(1,58) = 13.5$, $P < 0.005$; $F_2(1,20) = 11.9$, $P < 0.005$).

Experiment 3 confirmed that the duration of the ambiguous sequence, more than its lexical origin (i.e. excised from a monosyllabic word or the first syllable of a carrier word), influences its interpretation. Long sequences tended to be interpreted as mapping onto a monosyllabic word more than short sequences did. By selecting sequences from the same recording as in Experiment 1A on the basis of their duration, we were able to make the fixation pattern observed in Experiment 1A reverse. This confirms the importance of sequence duration in modulating the lexical interpretation of ambiguous sequences.

5. General discussion

This study examined the contribution of subphonemic, fine-grained acoustic cues to the activation of short words that occur at the onset of longer words, such as the monosyllabic word *ham* present at the onset of the carrier word *hamster*. Spliced carrier words

(e.g. *hamster*) were created by replacing the first syllable of an original recording of the carrier word with the recording of a monosyllabic word (e.g. *ham*) or with another token of the carrier word's first syllable. The effect of this manipulation on lexical access was evaluated by collecting participants' fixations to a picture representing the monosyllabic word (the competitor picture, e.g. the picture of a ham), as the spliced carrier word was heard. The proportion of fixations to the competitor picture was taken to reflect the degree of lexical activation of the monosyllabic word as the spliced carrier word was heard.

Experiment 1 showed that the competitor picture was fixated more when the first syllable of the spliced carrier word originated from a recording of the monosyllabic word than when it originated from another recording of the carrier word, revealing that the lexical interpretation of the ambiguous sequence (i.e. the first syllable of the spliced carrier word) was modulated by subphonemic acoustic cues. This demonstrates that the acoustic signal contained information that a purely phonemic description cannot capture. While this effect was found to be large and fully statistically reliable in Experiment 1A, where the monosyllabic word had been followed by a stressed syllable in its original recording, it was smaller and not fully significant in Experiment 1B, where the monosyllabic word had been followed by an unstressed syllable. Nevertheless, the statistically non-significant interaction between Experiments 1A and 1B suggests that the stress status of the following syllable does not have a reliable impact on the lexical interpretation of the ambiguous sequence. Rather, Experiment 1's results suggest that the disambiguation of an embedded sequence is subject to variability that the lexical origin of the embedded sequence could not account for.

Experiment 2 replicated Experiment 1A with different spliced stimuli. The spliced carrier words were created with tokens of the monosyllabic words and of the first syllable of the carrier words selected from our original recording with approximately equally short durations. The fixations to the competitor picture did not differ as a function of the origin of the ambiguous sequence of the spliced carrier word. In Experiment 3, the spliced carrier words were created with tokens of the monosyllabic words that were shorter than the tokens of the first syllable of the carrier words, in effect reversing the durational pattern of Experiment 1's stimuli. This time, the competitor picture was fixated more when the ambiguous sequence originated from the carrier word than when it originated from the monosyllabic word. Taken together, these results demonstrate that the duration of the ambiguous sequence, more than the word it originates from, determines its lexical interpretation.

The present study thus makes three important empirical contributions. First, it replicates the finding reported by Davis et al. (2002) with a different task, a different dependent measure, and a different language. Second, it extends it considerably by providing evidence that the production of a monosyllabic word or of the initial portion of a longer word does not always contain acoustic cues to disambiguation; which stimulus tokens were used affected the results. This possibility is rarely acknowledged in psycholinguistic research, where most often only one token per stimulus is tested. Third, this study contributes to our understanding of how the acoustic characteristics of embedded sequences can reduce lexical ambiguity by experimentally showing that the duration of the sequence, rather than its lexical origin, governs the degree to which lexical

candidates are considered. A long sequence tends to be interpreted as corresponding to a monosyllabic word more than a short sequence does.

These results have implications for accounts of speech production and for accounts of speech perception. We have argued that the differences between monosyllabic words and the first syllables of carrier words are a function of the prosodic structures that speakers build during the production of continuous speech. This claim is strongly supported by research in phonetics and phonology (as reviewed in Section 1), which has shown that prosodic boundaries influence the duration of preboundary segments. The prosodic-boundary hypothesis also provides a natural explanation for the variability that we have observed between productions of sentences with monosyllabic words and those with carrier words, and within the sets of each sentence type. Because the prosodic structure of an utterance is in part governed by factors that are independent of the morphosyntactic structure of the utterance, such as the speaker's speech rate, the production of a prosodic boundary after a monosyllabic word is not mandatory. Nevertheless, the acoustic correlates of a prosodic boundary are more likely to be associated with a monosyllabic word than with the first syllable of a polysyllabic word. As a result, a monosyllabic word tends to be of longer duration than the corresponding initial portion of a longer word, as was the case for the Davis et al. (2002) stimuli and for the Experiment 1 stimuli. Likewise, a prosodic boundary (and thus a longer word duration) was produced in our stimuli more often or more strongly when the monosyllabic word was followed by a stressed syllable than by an unstressed syllable, accounting for the robust effect of splicing in Experiment 1A and the inter-item variability observed in Experiment 1B.

In Section 1, we described an alternative account of the origin of these durational differences, namely, that they arise because they are lexically determined (i.e. durational information is specified as part of the lexical representation of words in the speech production system). Our results cast doubt on this account. It predicts that there should be two rather distinct sequence-duration distributions, depending on whether the sequence was produced as a monosyllabic word or as part of a longer word. Instead, we observed largely overlapping duration distributions. Furthermore, if durational information were lexically specified, the random selection of tokens in the monosyllabic-word conditions of Experiments 1A and 1B would have been made on the same duration distribution (i.e. that associated with monosyllabic words), predicting equivalent statistical outcomes on lexical disambiguation across these experiments, contrary to what we observed. Our results on the variability in surface realizations of sequence durations suggest that even if those durations were lexically specified, they would need to be adjusted post-lexically. The influence of prosodic structure on speech production could provide exactly that kind of post-lexical adjustment. Given the assumption that sequence duration is specified by prosodic structure, however, any prior lexical specification of duration appears to be redundant.

With regard to perception, we propose that the bias in interpreting an ambiguous sequence as a monosyllabic word, rather than a longer word, results from listeners predicting a prosodic boundary immediately following that sequence. We suggest 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. We thus take an integrated view of the production and perception of segmental variations in continuous speech, in which both

processes involve the computation of prosodic structure. It has been suggested that prosodic representations are computed as an utterance is processed, and that such representations contribute to processes such as the assignment of syntactic structure (e.g. Carlson, Clifton, & Frazier, 2001; Kjelgaard & Speer, 1999). If a prosodic structure has to be computed to contribute to establishing the syntactic structure of an utterance, it can also be used to modulate lexical activation.

According to our proposal, aspects of this prosodic structure, such as the edges of prosodic constituents equal to or higher than the word, could contribute to increasing the activation of lexical candidates whose boundaries are aligned with the hypothesized prosodic boundary. The effect of prosodic structure on lexical activation would operate in a probabilistic fashion so as to reflect the probabilistic relationship between segmental lengthening and the hypothesized word boundary. As demonstrated in the current study, a word boundary can occur after a sequence of a relatively short duration (Experiment 2) and segmental lengthening does not always coincide with a word boundary, presumably caused by other prosodic phenomena such as pitch accents (Experiment 3). Thus, the contribution of prosodic structure to lexical activation needs to be probabilistic. Furthermore, lexical information should be able to contribute to revising the prosodic structure if later-occurring segmental information most strongly supports a lexical hypothesis that is inconsistent with the hypothesized prosodic constituent.

Our pattern of results, however, is consistent with other accounts of lexical-embedding disambiguation. Exemplar models (e.g. Goldinger, 1998; Johnson, 1997a), for example, could in principle account for our results. In such models, fine-grained acoustic detail is represented in multiple lexical exemplars. The lexical representations of monosyllabic words could be characterized, among other things, by longer durations, and exemplars of carrier words could have shorter initial portions. This kind of model could thus explain the bias to interpret an ambiguous sequence as a monosyllabic word rather than as the initial part of a longer word when the acoustic realization of the sequence is longer: the more a token would match existing monosyllabic exemplars, the more likely it would be to be interpreted as a monosyllable. Johnson (1997b) provided simulations of an exemplar-based model that demonstrated such a bias. As the acoustic realization of the vocalic part of the word *cap* was presented to the model, the activation of the longer word *catalog* dropped while the activation of the words *cat* and *cap* remained high. The model was thus able to use the acoustic cues that were present in the tokens it had been trained on that distinguished monosyllabic words from longer words, and it was able to do so without explicitly encoding those cues in an abstract representation.

Another class of models that could potentially account for our results are those in which representations are more abstract than in exemplar models. Such models, including TRACE (McClelland & Elman, 1986), Shortlist (Norris, 1994) and the DCM (Gaskell & Marslen-Wilson, 1997) have abstract prelexical representations that recode the speech signal in some way prior to lexical access. In these models, fine-grained acoustic information could modulate lexical activation without the involvement of prosodic representations if it were encoded in prelexical representations and if the resulting activation of those representations were passed on to lexical representations.

The evidence presented here therefore does not demonstrate that lexical-embedding disambiguation is achieved via the computation of a prosodic structure by listeners. Attempts should be made to test this prosodic account against these alternative accounts. A challenge for any model is to specify exactly how fine-grained acoustic information, such as the segmental lengthening of ambiguous sequences, contributes to differential lexical activation. Regardless of how sequence duration influences lexical activation, it is most likely to be first analyzed in a context-dependent fashion. Variability in syllable durations in normal speech (e.g. as a function of speaking rate and style) is much greater than that in our experimental materials. Despite the fact that absolute sequence duration was a good predictor of the effects in the present study, this is unlikely to generalize across all types of utterance (e.g. the same absolute duration may be relatively long in one context and relatively short in another). Considerable work is therefore still required to establish how fine-grained acoustic details are used in a context-conditioned manner.

Finally, the exact nature of the acoustic cues that distinguish monosyllabic words from the initial portion of longer words needs to be established. The series of experiments presented here demonstrates that sequence duration is predictive of a bias in lexical interpretation. We used sequence duration as an index of the presence and/or strength of a prosodic boundary, based on the well-established effect of prosodic boundaries on preboundary segment duration (e.g. Beckman & Edwards, 1990; Turk & Shattuck-Hufnagel, 2000; Wightman et al., 1992). However, this in itself does not demonstrate that sequence duration is the dimension over which the computations leading to differential lexical activation take place. Segmental lengthening is likely to coincide with or trigger the realization of other acoustic cues, such as a larger pitch movement or degree of articulation. For example, in an analysis of linguopalatal contact in reiterant speech, Fougeron and Keating (1997) have shown that vowels are produced with greater articulatory magnitude in final position in the prosodic domain. Some or all of these acoustic cues may contribute to the postulation of a prosodic boundary, in proportion to the degree to which each cue is predictive of a word boundary.[5] Because segmental

[5] Measurements of the formant frequencies F1 and F2 on the sequences' vowels in the monosyllabic-word and carrier-word conditions in Experiment 1 evaluated the extent to which the context in which a sequence was produced (either as a monosyllabic word or as the first syllable of a longer word) affected the vowels' degree of articulation. In Experiment 1, analyses of the F1 and F2 values on the sequences' vowels indicated that the vowels' quality was affected by the context in which the sequence was produced. The vowel space, as defined by the averaged F1/F2 values for each of the nine different vowels found in the 28 experimental items, tended to be more expanded for sequences corresponding to monosyllabic words than for sequences found at the beginning of longer words. The expansion of the phonetic space was assessed by computing all 36 distances between the nine averaged vowels, and comparing the distances across conditions. Out of the 36 distances, 21 were larger in the monosyllabic-word stressed-context than in the carrier-word condition, and 23 were larger in the monosyllabic-word unstressed-context condition than in the carrier-word condition. However, simple sign tests established that this tendency was statistically unreliable ($P > 0.05$). The same analyses performed on the formant frequencies of the sequences' vowels in Experiments 2 and 3 showed differences in vowel space that were non-significant and, importantly, inconsistent with the tendency found in Experiment 1 or with the duration patterns manipulated in these experiments. These analyses, based on the admittedly very limited number of observations our stimuli offered, provided no reliable evidence that the vowels' articulation was consistently affected by the presence of a prosodic boundary.

lengthening strongly co-occurs with the presence of a word boundary, it is a good candidate for contributing to hypothesizing such a boundary. Moreover, the time course of some of the effects observed in the present experiments – weaker early in the ambiguous sequence than when the final part of the sequence was processed – is compatible with the view that the lexical interpretation of the sequence becomes increasingly biased toward a monosyllabic candidate as a long sequence unfolds over time. Nevertheless, the results of the current study do not directly speak to the issue of exactly which acoustic cues in the signal are used. Moreover, our cross-splicing manipulation involved the ambiguous sequence as well as the context that preceded it. The acoustic cues that contributed to the observed effects could have been located in the sequence itself, in its preceding context, or in both. Empirical tests involving the specific manipulation of the sequence's segmental duration are required to establish its direct role on lexical activation. Note, however, that if such experiments were to show that cues other than the sequence's duration (either in the ambiguous sequence or earlier) were in fact critical, such findings would not invalidate our more general suggestion that lexical activation is modulated by cues to prosodic structure.

The current study was motivated by the potential challenge that the pervasiveness of lexical embedding imposes on word-recognition models. The recognition of a word should be delayed until after its offset if this word is contained in a longer word. The current study has shown that the ambiguity resulting from lexical embedding is in fact not always as adverse as a phonemic transcription of the monosyllabic and carrier words would suggest, even in conditions where the ambiguity was maximized (by neutralizing semantic context and having the same phoneme(s) following the sequence). Although the presence of any bias is important in showing that the signal is encoded beyond the phonemes it contains, the strength of this bias was modest and the disfavored competitor remained active for a substantial amount of time after the disambiguating information was available. Davis et al. (2002) also found that the carrier-word interpretation was not ruled out until substantially after the disambiguating point (i.e. rejecting *captain* upon hearing "cap tucked"). These findings indicate that subtle acoustic cues resulting from segmental lengthening do not cause candidates to be ruled out. Instead, they appear to operate as a bias, favoring some alternatives over others.

As we pointed out when discussing Experiment 1A (Section 2.2.1), the time interval over which the fixations to the competitor picture remained high – after they started dropping – extended until quite late in time (between 800 and 900 ms in all experiments), later than what has been observed in past eye-tracking experiments examining the activation of cohort-like competitors, such as the activation of *beetle* when the target word *beaker* is heard (Allopenna et al., 1998; Dahan, Magnuson, & Tanenhaus, 2001; Dahan, Magnuson, Tanenhaus, & Hogan, 2001). Such a long interval was observed even when the ambiguous sequence originated from a carrier word. This suggests that the monosyllabic competitor remained in the competitor set for a substantial amount of time after bottom-up support for the carrier word was heard.

This long-lasting activation may have resulted from a number of factors. One obvious factor is the degree of activation the competitor reached before the information following the ambiguous sequence was heard and integrated. This activation level is likely to determine the time it takes for the competitor's activation to drop back to its resting level.

The degree of activation of a competitor is affected by the bottom-up support it receives (both in terms of strength and duration over time) and its lexical frequency. In addition, the competitor's activation may be modulated by competition with other activated words, such as the target word. From that perspective, the presentation of a target word at the end of an instruction such as "Click on the beaker" (as in Allopenna et al., 1998), where the segmentation of the target word from its right context is unproblematic, may result in stronger target activation and hence weaker competitor activation than when the target is embedded within a sentence, as in the present study. A more intriguing explanation for the long-lasting activation of the competitor, however, hinges on the fact that the information following the ambiguous sequence was not inconsistent with the monosyllabic-word interpretation until either it failed to match an existing word or it could not be parsed in a syntactically or semantically coherent manner. Competition associated with lexical embedding would thus take longer to resolve than the competition taking place between onset-overlapping words, such as *candy* and *candle*, where information that is inconsistent with the competitor is available as soon as the two words diverge. The existence of bottom-up inhibition (the use of inconsistent information to penalize mismatching words directly) is subject to debate, since inconsistent words can also be inhibited indirectly, via competition from matching words (see, e.g. Frauenfelder, Scholten, & Content, 2001). It will thus be important to determine whether the long-lasting activation of the monosyllabic competitors in the present study, compared to the activation of onset-overlapping competitors in other eye-tracking studies, provides evidence for bottom-up inhibition.

Our major finding, however, is that listeners can use the subphonemic acoustic cues often associated with the production of monosyllabic words, such as segmental lengthening, to bias their lexical interpretation of an utterance. This finding adds to a growing body of research that suggests that fine-grained subphonemic information in the speech signal can modulate lexical activation, both in the recognition of individual words (Andruski, Blumstein, & Burton, 1994; Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999) and in the recognition of words in continuous speech (Gow, 2002; Gow & Gordon, 1995; Spinelli, McQueen, & Cutler, 2003; Tabossi et al., 2000). Our results are also consistent with Davis et al. (2002), who showed that subphonemic cues can be used to resolve ambiguities caused by lexical embedding. We propose that the production of the acoustic cues that assist lexical disambiguation is not determined by properties that are inherent to the realization of monosyllabic or longer words, but depends on the realization of a prosodic boundary following monosyllabic words. We also propose that, in perception, the computation of a prosodic structure, built in parallel to the phonemic encoding of the signal, can affect lexical activation.

Acknowledgements

Part of this work was reported at the 42nd annual meeting of the Psychonomic Society, Orlando, FL, November 2001. We thank Mike Tanenhaus and Jim Magnuson for fruitful discussions in earlier stages of this work and three anonymous reviewers for helpful comments on this research.

Appendix A. Stimulus sets

Target	Competitor	Distractor	Distractor
beitel (chisel)	bij (bee)	vos (fox)	trechter (funnel)
bliksem (lightning)	blik (can)	hark (rake)	vissekom (fishbowl)
bokser (boxer)	bok (billy-goat)	peer (pear)	snijplank (chopping board)
cocktail (cocktail)	kok (chef)	tang (pliers)	schommel (swing)
compact-disc (CD)	kom (bowl)	bel (bell)	paprika (pepper)
eikel (acorn)	ei (egg)	bier (beer)	bureau (desk)
hamster (hamster)	ham (ham)	kraan (tap)	wasmachine (washing machine)
hendel (lever)	hen (hen)	loep (magnifier)	paperclip (paperclip)
kandelaar (candleholder)	kan (jug)	fee (fairy)	grasmaaier (lawn mower)
kijker (binoculars)	kei (stone)	vaas (vase)	molen (windmill)
knipsel (clipping)	knip (purse)	bas (bass)	vogelnest (bird's nest)
koekepan (frying pan)	koe (cow)	bril (glasses)	piramide (pyramid)
lama (llama)	la (drawer)	zaag (saw)	koptelefoon (headphones)
lampekap (lampshade)	lam (lamb)	web (web)	fornuis (stove)
leiding (pipe)	lei (slate)	hand (hand)	pompoen (pumpkin)
mantel (coat)	man (man)	boor (drill)	ladenkast (dresser)
panda (panda)	pan (pan)	bloes (shirt)	wekker (alarm clock)
panty (panty)	pen (pen)	mand (basket)	radijs (radish)
pinda (peanut)	pin (pin)	friet (fries)	ridder (knight)
regenton (rain barrel)	ree (deer)	haai (shark)	schoorsteen (chimney)
rooster (grid)	roos (rose)	been (leg)	vergiet (colander)
schilder (painter)	schil (peel)	tol (top)	microscoop (microscope)
slager (butcher)	sla (lettuce)	hoed (hat)	piano (piano)
snorkel (snorkel)	snor (moustache)	pijl (arrow)	waaier (fan)
taxi (taxi)	tak (branch)	berg (mountain)	helikopter (helicopter)
tegel (tile)	thee (tea)	kaas (cheese)	ananas (pineapple)
torso (torso)	tor (beetle)	slee (sleigh)	fakkel (torch)
zebra (zebra)	zee (sea)	stoel (chair)	fopspeen (pacifier)

Appendix B. Sentence sets

The first sentence in a sentence triplet corresponds to the carrier-word sentence that was presented in the experiments. The second and third sentences correspond to the sentences that mentioned the monosyllabic word in the stressed and unstressed contexts, respectively. Each sentence is followed by a phonetic transcription reflecting the speaker's realization of the carrier word or the monosyllabic word and its subsequent word.

Ik zag een BEITEL op de grond liggen.	ˈbɛi.təl
Ik zag een BIJ tussen de bloemen vliegen.	ˈbɛi ˈtʉ.sə
Ik zag een BIJ terugkeren naar de korf.	ˈbɛi tə.ˈʀʉχ.keː.ʀə
Ze zag een BLIKSEM in de verte.	ˈblɪk.səm
Ze zag een BLIK servicepaketten staan.	ˈblɪk ˈsʉʀ.vəs.pɑ.kɛ.tə
Ze zag een BLIK cement op tafel staan.	ˈblɪk səˈmɛnt
We wisten wel dat die oude BOKSER gestopt was.	ˈbɔk.səʀ
We wisten wel dat die oude BOK suffig was.	ˈbɔk ˈsʉ.fəχ
We wisten wel dat die oude BOK seniel was.	ˈbɔk sə.ˈniːl
Ik dacht dat die COCKTAIL het duurste was.	ˈkɔk.teːl
Ik dacht dat die KOK tekenlessen gaf.	ˈkɔk ˈteː.kən.lɛ.sə
Ik dacht dat die KOK tv-programma's maakte.	ˈkɔk teː.ˈveː.pʀoː.χʀɑ.maːs
Hij zei dat die COMPACT-DISC gevallen was.	ˈkɔm.pɑk.dɪsk
Hij zei dat die KOM pakjes bevatte.	ˈkɔm ˈpɑk.jəs
Hij zei dat die KOM pakketjes bevatte.	ˈkɔm pɑ.ˈkɛ.tjəs
Zij had een EIKEL gevonden.	ˈɛi.kəl
Zij had een EI kundig opgeverfd.	ˈɛi ˈkʉn.dəχ
Zij had een EI kunstmatig uitgebroed.	ˈɛi kʉnst.ˈmaː.təχ
Zij dacht dat die HAMSTER verdwenen was.	ˈhɑm.stəʀ
Zij dacht dat die HAM stukgesneden was.	ˈhɑm ˈstʉk.χə.sneː.də
Zij dacht dat die HAM steriel verpakt was.	ˈhɑm stə.ˈʀiːl
Hij zei dat die HENDEL niet meer functioneerde.	ˈhɛn.dəl
Hij zei dat die HEN duchtig met haar vleugels klapte.	ˈhɛn ˈdʉχ.təχ
Hij zei dat die HEN dezelfde was als daarstraks.	ˈhɛn də.ˈzɛlv.də
Ik geloof dat die KANDELAAR er niet meer is.	ˈkɑn.də.laːʀ
Ik geloof dat die KAN dubbel zo veel kostte.	ˈkɑn ˈdʉ.bəl
Ik geloof dat die KAN dezelfde kleur heeft.	ˈkɑn də.ˈzɛlv.də
Hij had die KIJKER meegenomen.	ˈkɛi.kəʀ
Hij had die KEI kundig ingepakt.	ˈkɛi ˈkʉn.dəχ
Hij had die KEI kunstzinnig beschilderd.	ˈkɛi kʉnst.ˈsɪ.nəχ

Ze probeerde haar KNIPSEL op te zoeken.	ˈknɪp.səl
Ze probeerde haar KNIP sullig dicht te maken.	ˈknɪp ˈsʉ.ləχ
Ze probeerde haar KNIP secuur te sluiten.	ˈknɪp sə.ˈkyːʀ
Hij dacht dat die KOEKEPAN van hem was.	ˈkuː.kə.pɑn
Hij dacht dat die KOE kuddedieren meed.	ˈkuː ˈkʉ.də.diː.ʀə
Hij dacht dat die KOE cultuurgewas luste.	ˈkuː kʉl.ˈtyːʀ.χə.ʊɑs
Met die LAMA is niets aan de hand geweest.	ˈlaː.maː
Met die LA maatdoppen kun je aan de slag.	ˈlaː ˈmaː.dɔ.pə
Met die LA manuscripten kun je aan de slag.	ˈlaː maː.nʉ.ˈskʀɪp.tə
Hij zei dat een LAMPEKAP aangeschaft was.	ˈlɑm.pə.kɑp
Hij zei dat een LAM pudding mocht eten.	ˈlɑm ˈpʉ.dɪŋ
Hij zei dat een LAM personen zou mijden.	ˈlɑm pəʀ.ˈsoː.nə
Ze zag dat de LEIDING er niet meer was.	ˈlɛi.dɪŋ
Ze zag dat de LEI dichtgeklapt was.	ˈlɛi ˈdɪχt.χə.klɑpt
Ze zag dat de LEI discreet verstopt was.	ˈlɛi dɪs.ˈkʀeːt
Hij probeerde de MANTEL te verkopen.	ˈmɑn.təl
Hij probeerde de MAN tussentijds te helpen.	ˈmɑn ˈtʉ.sə.tɛits
Hij probeerde de MAN tegemoet te lopen.	ˈmɑn tə.χə.ˈmuːt
Ik zag dat de PANDA er niet meer was.	ˈpɑn.daː
Ik zag dat de PAN dadels bevatte.	ˈpɑn ˈdaː.dəls
Ik zag dat de PAN daarachter gezet was.	ˈpɑn daːʀ.ˈɑχ.təʀ
Ik vond dat die PANTY haar niet zo goed stond.	ˈpɛn.tiː
Ik vond dat die PEN typisch gevormd was.	ˈpɛn ˈtiː.piːs
Ik vond dat die PEN timide schreef.	ˈpɛn tiː.ˈmiː.də
Ik wilde de PINDA opeten.	ˈpɪn.daː
Ik wilde de PIN daarom vast prikken.	ˈpɪn ˈdaːʀ.ɔm
Ik wilde de PIN daarachter steken.	ˈpɪn daːʀ.ˈɑχ.təʀ
Hij vertelde dat die REGENTON daar niet meer stond.	ˈʀeː.χən.tɔn
Hij vertelde dat die REE gulzig van aard was.	ˈʀeː ˈχʉl.zəχ
Hij vertelde dat die REE genoeg gegeten had.	ˈʀeː χə.ˈnuːχ

Zij had een ROOSTER van me meegekregen.	ˈʀoːs.təʀ
Zij had een ROOS tussen het boeket gestopt.	ˈʀoːs ˈtʉ.sə
Zij had een ROOS teveel aan hem verkocht.	ˈʀoːs tə.ˈveːl
Zij dacht dat die SCHILDER hem had geholpen.	ˈsχɪl.dəʀ
Zij dacht dat die SCHIL dubbelgevouwen was.	ˈsχɪl ˈdʉ.bəl.χə.vɑu.ʋə
Zij dacht dat die SCHIL dezelfde vorm zou hebben.	ˈsχɪl dəˈzɛlv.də
Je mag die SLAGER daar de schuld van geven.	ˈslaː.χəʀ
Je mag die SLA gulzig gaan opeten.	ˈslaː ˈχʉl.zəχ
Je mag die SLA gerust even schoonmaken.	ˈslaː χəˈʀʉst
Hij zei dat die SNORKEL niet van hem was.	ˈsnɔʀ.kəl
Hij zei dat die SNOR kunstig versierd was.	ˈsnɔʀ ˈkʉn.stəχ
Hij zei dat die SNOR kunstmatig verlengd was.	ˈsnɔʀ kʉnstˈmaː.təχ
Ze probeerde de TAXI in het zicht te houden.	ˈtɑk.siː
Ze probeerde de TAK sinaasappels te pakken.	ˈtɑk ˈsiː.naːs.ɑ.pəls
Ze probeerde de TAK citroenen te pakken.	ˈtɑk siː.ˈtʀuː.nə
Ik kon de TEGEL zonder veel moeite pakken.	ˈteː.χəl
Ik kon de THEE gulzig gaan opdrinken.	ˈteː ˈχʉl.zəχ
Ik kon de THEE gelukkig nog ruilen.	ˈteː χəˈlʉ.kəχ
Hij probeerde een TORSO uit elkaar te halen.	ˈtɔʀ.zoː
Hij probeerde een TOR zomaar op te pakken.	ˈtɔʀ ˈzoː.maːʀ
Hij probeerde een TOR zolang op te bergen.	ˈtɔʀ zoː.ˈlɑŋ
Hij vertelde dat de ZEBRA ontsnapt was.	ˈzeː.bʀaː
Hij vertelde dat de ZEE brasems bevat.	ˈzeː ˈbʀaː.səms
Hij vertelde dat de ZEE Brazilië omringt.	ˈzeː bʀaː.ˈziː.liː.jə

References

Abercrombie, D. (1967). *Elements of general phonetics*. Edinburgh: Edinburgh University Press.

Allopenna, P. D., Magnuson, J. S., & Tanenhaus, M. K. (1998). Tracking the time course of spoken word recognition using eye movements: evidence for continuous mapping models. *Journal of Memory and Language*, *38*, 419–439.

Andruski, J. E., Blumstein, S. E., & Burton, M. (1994). The effect of subphonetic differences on lexical access. *Cognition*, *52*, 163–187.

Baayen, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database (CD-ROM)*. Philadelphia, PA: Linguistic Data Consortium, University of Pennsylvania.

Bard, E. G., Shillcock, R. C., & Altmann, G. T. M. (1988). The recognition of words after their acoustic offsets in spontaneous speech: effects of subsequent context. *Perception & Psychophysics*, *44*, 395–408.

Beckman, M. E., & Edwards, J. (1990). Lengthenings and shortenings and the nature of prosodic constituency. In J. Kingston, & M. E. Beckman (Eds.), *Papers in laboratory phonology. I. Between the grammar and the physics of speech* (pp. 152–178). Cambridge: Cambridge University Press.

Beckman, M. E., & Pierrehumbert, J. B. (1986). Intonational structure in Japanese and English. *Phonology Yearbook*, *3*, 255–309.

Cambier-Langeveld, T. (2000). *Temporal marking of accents and boundaries*. Leiden: Holland Institute of Generative Linguistics.

Carlson, K., Clifton, C., Jr., & Frazier, L. (2001). Prosodic boundaries in adjunct attachment. *Journal of Memory and Language*, *45*, 58–81.

Christophe, A., Dupoux, E., Bertoncini, J., & Mehler, J. (1994). Do infants perceive word boundaries? An empirical study of the bootstrapping of lexical acquisition. *Journal of the Acoustical Society of America*, *95*, 1570–1580.

Christophe, A., Mehler, J., & Sebastián-Gallés, N. (2001). Perception of prosodic boundary correlates by newborn infants. *Infancy*, *2*, 385–394.

Cooper, W. E., & Paccia-Cooper, J. (1980). *Syntax and speech*. Cambridge, MA: Harvard University Press.

Cycowicz, Y. M., Friedman, D., Rothstein, M., & Snodgrass, J. G. (1997). Picture naming by young children: norms for name agreement, familiarity, and visual complexity. *Journal of Experimental Child Psychology*, *65*, 171–237.

Dahan, D., Magnuson, J. S., & Tanenhaus, M. K. (2001). Time course of frequency effects in spoken-word recognition: evidence from eye movements. *Cognitive Psychology*, *42*, 317–367.

Dahan, D., Magnuson, J. S., Tanenhaus, M. K., & Hogan, E. M. (2001). Subcategorical mismatches and the time course of lexical access: evidence for lexical competition. *Language and Cognitive Processes*, *16*, 507–534.

Davis, M. H., Gaskell, M. G., & Marslen-Wilson, W. (1997). Recognising embedded words in connected speech: context and competition. In J. A. Bullinaria, D. W. Glasspool, & G. Houghton (Eds.). *Proceedings of the fourth neural computation and psychology workshop: connectionist representations*, (pp. 254–266). London: Springer-Verlag.

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.

Elman, J. L. (1990). Finding structure in time. *Cognitive Science*, *14*, 179–211.

Fougeron, C., & Keating, P. A. (1997). Articulatory strengthening at edges of prosodic domains. *Journal of the Acoustical Society of America*, *101*, 3728–3740.

Frauenfelder, U. H. (1991). Lexical alignment and activation in spoken word recognition. In J. Sundberg, L. Nord, & R. Carlson (Eds.), *Music, language, speech and brain* (pp. 294–303). *Wenner-Gren International Symposium Series*, London: Macmillan.

Frauenfelder, U. H., & Peeters, G. (1990). Lexical segmentation in TRACE: an exercise in simulation. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives* (pp. 50–86). Cambridge, MA: MIT Press.

Frauenfelder, U. H., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, *16*, 583–607.

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.

Gaskell, M. G., & Marslen-Wilson, W. D. (1999). Ambiguity, competition, and blending in spoken word recognition. *Cognitive Science*, *23*, 439–462.

Gee, J. P., & Grosjean, F. (1983). Performance structures: a psycholinguistic and linguistic appraisal. *Cognitive Psychology*, *15*, 411–458.

Goldinger, S. D. (1998). Echoes of echoes? An episodic theory of lexical access. *Psychological Review*, *105*, 251–279.

Gow, D. W., Jr. (2002). Does English coronal place assimilation create lexical ambiguity? *Journal of Experimental Psychology: Human Perception and Performance*, *28*, 163–179.

Gow, D. W., Jr., & 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.

Grosjean, F. (1985). The recognition of words after their acoustic offset: evidence and implications. *Perception & Psychophysics*, *38*, 299–310.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10.1–10.112). New York: Wiley.

Harris, M. S., & Umeda, N. (1974). Effect of speaking mode on temporal factors in speech: vowel duration. *Journal of the Acoustical Society of America*, *56*, 1016–1018.

Johnson, K. (1997a). Speech perception without speaker normalization: an exemplar model. In K. Johnson, & J. W. Mullennix (Eds.), *Talker variability in speech processing* (pp. 145–165). San Diego, CA: Academic Press.

Johnson, K. (1997b). The auditory/perceptual basis for speech segmentation. *Ohio State University Working Papers in Linguistics*, *50*, 101–113.

Jones, D. (1972). *An outline of English phonetics*. Cambridge: Cambridge University Press.

Kjelgaard, M. M., & Speer, S. R. (1999). Prosodic facilitation and interference in the resolution of temporary syntactic closure ambiguity. *Journal of Memory and Language*, *40*, 153–194.

Klatt, D. H. (1976). Linguistic uses of segmental duration in English: acoustic and perceptual evidence. *Journal of the Acoustical Society of America*, *59*, 1208–1221.

Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: evidence from syllable duration. *Proceedings of the XIIth International Congress of Phonetic Sciences*, (pp. 290–293). Aix-en-Provence: University of Provence.

Lehiste, I. (1972). The timing of utterances and linguistic boundaries. *Journal of the Acoustical Society of America*, *51*, 2018–2024.

Liberman, M., & Prince, A. (1977). On stress and linguistic rhythm. *Linguistic Inquiry*, *8*, 249–336.

Luce, P. A (1986a). Neighborhoods of words in the mental lexicon (PhD dissertation, Indiana University). In: Research on speech perception, Technical Report No. 6, Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P. A. (1986b). A computational analysis of uniqueness points in auditory word recognition. *Perception & Psychophysics*, *39*, 155–158.

Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.

Marslen-Wilson, W., & Warren, P. (1994). Levels of perceptual representation and process in lexical access: words, phonemes, and features. *Psychological Review*, *101*, 653–675.

Martin, J. G. (1970). On judging pauses in spontaneous speech. *Journal of Verbal Learning and Verbal Behavior*, *9*, 75–78.

McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *18*, 1–86.

McQueen, J. M., Cutler, A., Briscoe, T., & Norris, D. (1995). Models of continuous speech recognition and the contents of the vocabulary. *Language and Cognitive Processes*, *10*, 309–331.

McQueen, J. M., Norris, D., & Cutler, A. (1994). Competition in spoken word recognition: spotting words in other words. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *20*, 621–638.

McQueen, J. M., Norris, D., & Cutler, A. (1999). Lexical influence in phonetic decision making: evidence from subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance*, *25*, 1363–1389.

Nespor, M., & Vogel, I. (1986). *Prosodic phonology*. Dordrecht: Foris.

Nooteboom, S. G., & Doodeman, G. J. N. (1980). Production and perception of vowel length in spoken sentences. *Journal of the Acoustical Society of America*, *67*, 276–287.

Norris, D. (1990). A dynamic-net model of human speech recognition. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives* (pp. 87–104). Cambridge, MA: MIT Press.

Norris, D. (1994). Shortlist: a connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.

Oller, D. K. (1973). The effect of position in utterance on speech segment duration in English. *Journal of the Acoustical Society of America*, *54*, 1235–1247.

Pierrehumbert, J., & Liberman, M. (1982). Modeling the fundamental frequency of the voice (review of Cooper & Sorensen, 1981). *Contemporary Psychology*, *27*, 690–692.

Quené, H. (1992). Durational cues for word segmentation in Dutch. *Journal of Phonetics*, *20*, 331–350.

Rakerd, B., Sennett, W., & Fowler, C. A. (1987). Domain-final lengthening and foot-level shortening in spoken English. *Phonetica*, *44*, 147–155.

Selkirk, E. O. (1984). *Phonology and syntax: the relation between sound and structure*. Cambridge, MA: MIT Press.

Shattuck-Hufnagel, S., & Turk, A. E. (1996). A prosody tutorial for investigators of auditory sentence processing. *Journal of Psycholinguistic Research*, *25*, 193–247.

Smits, R., Warner, N., McQueen, J. M., & Cutler, A. (2003). Unfolding of phonetic information over time: a database of Dutch diphone perception. *Journal of the Acoustical Society of America*, *113*, 563–574.

Snodgrass, J. G., & Vanderwart, M. (1980). A standardized set of 260 pictures: norms for name agreement, image agreement, familiarity, and visual complexity. *Journal of Experimental Psychology: Human Learning and Memory*, *6*, 174–215.

Spinelli, E., McQueen, J. M., & Cutler, A. (2003). Processing resyllabified words in French. *Journal of Memory and Language*, *48*, 233–254.

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.

Turk, A. E., & Shattuck-Hufnagel, S. (2000). Word-boundary-related duration patterns in English. *Journal of Phonetics*, *28*, 397–440.

Vroomen, J., & de Gelder, B. (1997). Activation of embedded words in spoken word recognition. *Journal of Experimental Psychology: Human Perception and Performance*, *23*, 710–720.

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.

Zwitserlood, P. (1989). The locus of the effects of sentential-semantic context in spoken-word processing. *Cognition*, *32*, 25–64.

Zwitserlood, P., & Schriefers, H. (1995). Effects of sensory information and processing time in spoken-word recognition. *Language and Cognitive Processes*, *10*, 121–136.