PROSODICALLY-CONDITIONED DETAIL IN THE RECOGNITION OF SPOKEN WORDS

© 2005, Anne Pier Salverda
Cover design: Linda van den Akker & Inge Doehring
Cover illustration: Alexandra Crouwers
Printed and bound by Ponsen & Looijen bv, Wageningen
ISBN 90-76203-20-2

PROSODICALLY-CONDITIONED DETAIL IN THE RECOGNITION OF SPOKEN WORDS

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor
aan de Radboud Universiteit Nijmegen
op gezag van de Rector Magnificus prof. dr. C.W.P.M. Blom,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 19 december 2005
des namiddags om 1.30 uur precies

door
Anne Pier Salverda
geboren op 26 april 1975
te Wageningen

Promotor:	Prof. dr. A. Cutler
Copromotores:	Dr. D. Dahan (MPI en University of Pennsylvania) Dr. J.M. McQueen (MPI)
Manuscriptcommissie:	Prof. dr. W. Vonk Prof. dr. P. Zwitserlood (Westfälische Wilhelms-Universität Münster) Dr. H. Strik

The research reported in this thesis was supported by a grant from the Max-Planck-Gesellschaft zur Förderung der Wissenschaften, München, Germany.

The eye sees only what the mind is prepared to comprehend

Henri Bergson (1859—1941)

VOORWOORD

Zo voelt het dus om—voor heel even—eindelijk met beide benen op de horizon te staan. Toen ik in 1999 als onderzoeksassistent het Max Planck Instituut (MPI) kwam binnenrollen, had ik echt niet durven denken dat ik er op een goede dag met een proefschrift uit zou rollen.

Heel veel dank ben ik verschuldigd aan mijn begeleiders, Delphine Dahan en James McQueen. Betere copromotores had ik me niet kunnen wensen. Delphine heeft met haar nimmer aflatende inzet, energie, aandacht, vertrouwen en geduld wat mij betreft een *gold standard* gezet voor het begeleiden van promovendi. James ben ik bijzonder dankbaar voor zijn steun, altijd op het juiste moment, tijdens de laatste maanden waarin ik aan mijn proefschrift werkte. Mijn promotor Anne Cutler wil ik bedanken voor de inspirerende en stimulerende wetenschappelijke omgeving die zij weet te creëren en onderhouden in de *Comprehension Group* van het MPI. Iedereen die in de jaren van mijn verblijf een bijdrage heeft geleverd aan de bijzondere sfeer in deze groep wil ik bedanken, met name mijn kamergenoten Mirjam Broersma, Christiane Dietrich, Martijn Goudbeek, Claudia Kuzla en Keren Shatzman.

Naast een bron van wetenschappelijk welbehagen was het MPI ook in alledaags opzicht een omgeving waarin ik me altijd bijzonder heb thuisgevoeld. De opgewekte stemming van Agnes, José, Hans, Jan, John, Kees, Rian en Pim-van-de-kantine zorgde er voor dat mijn dag meestal al geslaagd was voordat ik goed en wel achter mijn bureau had plaatsgenomen. Het MPI volleybal team hield me scherp en in conditie.

In 2002 bracht ik, op uitnodiging van Mike Tanenhaus, een jaar door in het *Department of Brain and Cognitive Sciences* van de University of Rochester. Ik had in Rochester de tijd van mijn leven, mede dankzij Mike's ongeëvenaarde gastvrijheid en de unieke en levendige atmosfeer in zijn lab. In het bijzonder wil ik mijn vrienden Sarah Brown-Schmidt, Ellen Campana, Marie Coppola en Jessica Maye bedanken voor hun bijdragen aan mijn Amerikaanse avontuur. Mijn kamergenoten Dan Koo en Mrim Boutla, die mij *American Sign Language* leerden, zorgden er voor dat het me nauwelijks opviel dat er geen ramen in ons kantoor zaten.

In Nijmegen zorgden mijn goede vrienden Jan van de Mortel en Oliver Müller voor de nodige afleiding in de vorm van diepzinnige en onzinnige gesprekken, culinaire avonturen, muziek, boeken en filmbezoek, zodat ik me ook buiten werktijd altijd goed vermaakte.

Gerry Altmann en Mike Tanenhaus ben ik erkentelijk voor het feit dat zij, tijdens mijn pre-doc respectievelijk post-doc, af en toe een oogje dichtknepen wanneer er wat werk aan mijn proefschrift moest worden verricht.

Tenslotte een bijzonder woord van dank voor mijn ouders Netty en Marten, mijn broertjes Joost en Merijn en mijn zusje Irene, die ieder op hun eigen manier altijd voor me klaarstonden met hun vertrouwen en steun.

TABLE OF CONTENTS

CHAPTER 1

CHAPTER 2

INTRODUCTION

CHAPTER 1

A substantial amount of our daily lives is dedicated to the use and processing of spoken language. A speaker, by modifying the flow of air through his vocal tract as he exhales, creates a physical signal consisting of air molecules in motion, which can convey an infinite number of different messages to a listener. When this signal reaches the listener's ears, she can start the process of reconstructing the message of the speaker on the basis of her resonating eardrums. In everyday life, this extraordinary achievement is easily taken for granted, since understanding spoken language usually requires very little effort. It is only when confronted with a foreign language which one does not speak or understand that suddenly the listener can have the disconcerting experience that making sense of the speech signal is a far from trivial process.

In order to understand the message of the speaker, a listener must retrieve the meaning of words in the speech signal. This requires the listener to recognize the words that the speaker produces. A prerequisite for this ability is therefore to know what each of the words of the ambient language sounds like. Such information is stored in the listener's mental lexicon and associated with information about the meaning of each of those words. Spoken language comprehension entails extracting all the relevant information from the speech signal and subsequently mapping this information onto the sound-form representations of words in the mental lexicon. The research that is presented in this thesis is concerned with this mapping process and, in particular, with the information in the speech signal that is relevant for lexical processing. Several characteristics of the speech signal impose important constraints on the processes that are involved in the recognition of spoken words.

Speech is temporal

First of all, spoken language unfolds over time. Speech is therefore a rapidly changing temporal signal and an important characteristic of the acoustic events in this signal is that they are transient in nature. In order to process the information in the speech signal effectively, a listener must therefore rapidly analyze the speech signal. Many studies have demonstrated that listeners have this ability and that the processing of spoken language is very closely time locked to the input (Marslen-Wilson, 1973, 1975; Zwit-

serlood, 1989). As the speech signal unfolds, the recognition of spoken words proceeds in an incremental fashion. As soon as acoustic information becomes available, this information is used to develop and evaluate hypotheses about the interpretation of the speech signal. At any point in time, words that are consistent with the speech input are activated in parallel and compete for recognition. Because the initial sounds of a word are usually consistent with multiple lexical interpretations, the recognition of a spoken word is essentially a process of ambiguity resolution. The initial sounds of the word *candy*, /kæ../, are consistent with the words *candy*, *captain* and *canvas*, and a listener will thus have to consider these candidate words for recognition. As more acoustic information becomes available, e.g. /kæn../, lexical hypotheses that are inconsistent with this information (e.g., *captain*) can be ruled out. A spoken word can be uniquely identified as soon as the speech signal is consistent with only one candidate.

Speech is continuous

A second important characteristic of the speech signal is that it is largely continuous. The listener's subjective impression is that speech consists of a sequence of discrete words, but one look at a physical representation of the signal reveals that this signal does not contain an equivalent of the blank spaces between words that occur in a printed text. The recognition of spoken words therefore also entails the segmentation of the speech signal into discrete words. Some boundaries between words in continuous speech are marked. For instance, in English, word-initial voiceless stops are aspirated, and the initial and final segments of words may be lengthened. However, such cues are often small and variable, and because the presence of such information in the speech signal is not reliable, listeners cannot rely exclusively on these cues to word boundaries to solve the segmentation problem.

Speech is variable

Third, the speech signal is highly variable. To characterize a spoken word as a sequence of speech sounds does not do justice to the variability that is encountered between different realizations of the same word. The acoustic realization of a word can be affected by many factors. Some of these factors are idiosyncracies of the speaker, e.g. age, gender, dialect and speaking rate. However, even different tokens of a particular word produced by the same speaker will never be realized in exactly the same way. This is because the realization of a word also depends on its segmental context. The realization of a speech sound is in part influenced by the realization of its neighboring speech sounds, a phenomenon that is called coarticulation. The realization of

the initial sound of the word *candy* is therefore different in the phrase *more candy* than it is in the phrase *less candy*. Moreover, even when a speaker produces the phrase *more candy* several times, the word *candy* will never be pronounced in exactly the same way. The listener's ability to identify words successfully despite the fact that a spoken word can be realized in an infinite number of different ways constitutes a remarkable human cognitive skill.

PRELEXICAL REPRESENTATIONS

Despite all the variability associated with the realization of a word across different contexts, and even variability associated with different realizations of a word within the same context, listeners have the ability to recognize each of those tokens. It is therefore generally assumed that the mapping between the auditory speech signal and sound-form representations of words in the mental lexicon is mediated by prelexical representations that abstract away from variability associated with different realizations of a word. Current theories and models of spoken-word recognition that incorporate prelexical representations make different assumptions about the nature of these representations. Importantly, the nature of such representations constrains the recognition process, since only information that is preserved in these representations can affect the mapping of the speech signal onto lexical representations. Prelexical representations therefore reflect a theory's or a model's claims regarding the information in the speech signal that is relevant to distinguish a word from other words. In other words: information that is not captured by prelexical representations is assumed to be irrelevant for spoken-word recognition.

MODELS OF SPOKEN-WORD RECOGNITION

The Cohort model

The sound form of a word is often described as a sequence of phonemes, which correspond to the smallest contrastive units in the sound system of a language. Each of the words of a language, when described as a sequence of phonemes, is therefore associated with a unique phonemic representation[1]. Such representations can thus capture differences in sound forms between all the words of a language, abstracting away from acoustic-phonetic information in the speech signal that may not be relevant for

[1] Except for homophones: words that differ in meaning but that are pronounced alike, e.g. "sea" and "see".

making lexical distinctions. This phonemic perspective was shared by many early models of spoken-word recognition, such as the Cohort model (Marslen-Wilson & Welsh, 1978), which assumed that the representations that are involved in the recognition of spoken words are phonemic in nature. (Although Marslen-Wilson and Welsh (1978) do not state explicily that the Cohort model relies on phonemic representations, it is reasonable to make this assumption, given their description of how the model evaluates candidate words.) That is, in this model, both the speech signal and the sound-form representations of words in the mental lexicon are represented as sequences of phonemes. A phonemic representation of the speech signal, which is built as the signal is processed, is continuously matched against the sound-form representations of words in the mental lexicon. At any point in time, words that are consistent with the speech signal are activated. For instance, upon hearing the spoken sequence /kæ../, all words that start with these sounds, such as *candy*, *captain* and *canvas* are activated in parallel, comprising a cohort of candidate words that are considered for recognition. As more information about the sound form of the word becomes available (e.g., /kæn../), candidate words that mismatch this information (e.g., *captain*) are deactivated and the size of the cohort is reduced. Recognition occurs when the cohort constitutes of a single word, and the recognition process starts again for the next word. Because the sound form of a word can diverge from that of all other words before its offset, an important prediction of the Cohort model, which has been confirmed in numerous studies, is that word recognition can often be achieved before the entire word has been heard. The point at which a word's sound form diverges from that of all other words is called its uniqueness point. For instance, as soon as the listener has heard the sounds /kəuhɔ../, this sequence can be attributed to the word *cohort*, which is the only word in the lexicon that starts with these sounds.

The major contribution of the Cohort model is that it provided a basic framework for spoken-word recognition by distinguishing three important processes: an activation process, an evaluation process and a selection process. The interplay of these processes results in an optimal process of ambiguity resolution in which the recognition of a spoken word proceeds in an incremental fashion. However, the architecture of the Cohort model renders the identification of words in the speech signal a strictly sequential process. This is because the model relies on the successful identification of a word to predict where the next word will start, in order to allow for the recognition process to start again. In other words: the model relies on the successful identification of word boundaries in order for the recognition process to proceed. In continuous speech, however, the locations of word boundaries are often not available to the listener. A major challenge for the Cohort model's use of lexical information to

locate word boundaries was put forward by Luce (1986). He showed that in English, many monosyllabic words do not become unique before their offset. This means that a large proportion of English words, and especially monosyllabic words of high frequency, are embedded at the onset of longer words (e.g., the word *can*, which phonemically matches the initial sounds of longer words such as *candy* and *candle*). The prevalence of word-initial embedding in the vocabulary renders the Cohort model's anticipation strategy problematic, since many words can in fact not be uniquely identified before their offset. For instance, the spoken sequence /kændi../ cannot unambiguously be attributed to the word *candy*, because it can also correspond to the word *can* followed by a word starting with the sounds /di/. The fundamental limitation of the Cohort model is that although the identification of the location of the onsets of words is a prerequisite for successful recognition, the model's mechanisms appear inadequate to reliably locate such information.

In a more recent and improved version of the Cohort model (Marslen-Wilson, 1987), lexical candidates are evaluated more flexibly than in the original Cohort model, where a candidate word was either part of the cohort of activated candidates or not. This was achieved by modifying the model's architecture and its representations. First, in the 1987 version of the Cohort model, prelexical representations are assumed to be featural instead of phonemic. This allows for a more fine-grained evaluation of lexical candidates, rendering this process sensitive to subphonemic information. Second, the degree of support for lexical candidates, as reflected by their lexical activation, is computed in a graded fashion. Thus, candidate words that are considered for recognition are activated in parallel, each to the degree that they are supported by information in the speech signal, so that at each moment during the processing of the unfolding speech signal, some candidate words are considered for recognition more strongly than others.

TRACE

The TRACE model of speech perception (McClelland & Elman, 1986) was the first computational model of spoken-word recognition. This connectionist model consists of an interactive-activation network that distinguishes three separate but interconnected levels of processing: a featural, a phonemic and a word level. Nodes within each level correspond to perceptual hypotheses, and the activation of each node in the network reflects the degree of support for those hypotheses. Activation in the network spreads because nodes excite connected nodes that converge on the same hypotheses at adjacent levels of processing. Nodes within each level of processing represent different hypotheses and therefore inhibit each other. In TRACE, the entire network of

interconnnected nodes is represented at every successive time slice. The input to the model consists of a featural representation of the speech input, which activates feature nodes. These nodes in turn activate phonemic nodes, which then activate word nodes. In contrast to the Cohort model, TRACE does not give priority to information associated with the onset of words, and candidate words can thus be activated by any part of the speech signal. Recognition occurs when a word node reaches a certain threshold of activation.

TRACE improved on the 1978 version of the Cohort model in two fundamental ways. First, the model's architecture and featural representations render lexical activation a graded process, with each word being activated in proportion to the support that it receives from the speech signal. The evaluation of candidate words in TRACE is thus different in nature from the evaluation of candidate words in the Cohort model, where a word was either part of the cohort of activated words or not. For instance, TRACE predicts that the initial sounds of the word *bear* will not only activate the candidate *bear* but also the candidate *pear*, because of the featural overlap between the phonemes /b/ and /p/. Because lexical activation is a function of the degree to which a candidate word matches the speech signal, the model can recognize slightly mispronounced words, such as *shigarette*, because this word will most strongly activate the candidate *cigarette*. Second, in TRACE, all candidate words compete with each other, and the degree of activation of candidates is thus influenced by the degree of activation of other candidates. This lexical competition process acts to increase initial differences in activation levels between candidates that arise as a function of their initial goodness of fit with the speech signal. The lexical competition process ensures that an optimal parse of the speech input is achieved. Word boundaries are not identified but simply emerge as a result of the competition process, even when no word boundaries are marked in the input.

TRACE can successfully simulate a broad range of experimental findings on the recognition of spoken words, such as competition between simultaneously activated candidate words (McQueen, Norris, & Cutler, 1994), the activation of words that are embedded in longer words (Gow & Gordon, 1995; Shillcock, 1990; Vroomen & de Gelder, 1997), the activation of words that straddle word boundaries (Tabossi, Burani, & Scott, 1995), and the influence of subphonemic variation on lexical activation (Andruski, Blumstein, & Burton, 1994). However, the model cannot account for the findings of several studies that suggest that listeners rely on explicit segmentation strategies to assist the recognition process. For instance, in English, metrical information and phonotactic information can influence the lexical activation of candidate words (Cutler & Norris, 1988; McQueen, 1998).

Shortlist

Shortlist (Norris, 1994; Norris, McQueen, Cutler, & Butterfield, 1997), like TRACE, is a competition-based connectionist model of spoken-word recognition. The input to the model consists of a sequence of phonemes that constitutes a phonemic analysis of the speech signal. Candidate words can be activated by any portion of the speech signal, and the activation of each candidate word is determined by the degree to which it matches and mismatches the speech signal. A set of candidate words that consists of only those candidates that are strongly supported by the input is subsequently generated and wired into a small interactive network. In this network, candidates compete with each other for recognition, and segmentation of the input is thus achieved in a similar way to TRACE.

An important difference between the most recent version of Shortlist (Norris et al., 1997) and TRACE is that Shortlist develops hypotheses about the location of word boundaries in the speech input. This information is used to improve the lexical competition process by reducing the activation levels of candidate words that are misaligned with hypothesized word boundaries. For instance, because the sound sequence /fn/ is an illegal onset cluster in English, this sequence signals the location of a likely word boundary between the /f/ and the /n/. The present version of Shortlist uses metrical and phonotactic information as cues to likely word boundaries, positing word boundaries at the onset of strong syllables (because most words in the English language start with a strong syllable) and between two sounds that do not co-occur within a syllable (e.g., /fn/).

Direct-mapping models

A class of models that is radically different from the ones discussed so far is the class of exemplar and episodic models of spoken-word recognition (e.g., Goldinger, 1998; Johnson, 1997; Klatt, 1979). An important and defining characteristic of these so-called direct-mapping models is that the mapping of the speech input onto stored representations of lexical form is not mediated by prelexical representations. Instead, the speech signal is mapped directly onto lexical representations. These representations therefore include a large amount of phonetic detail associated with the realization of a spoken word. Sound-form representations of words are acquired on the basis of experience with the ambient language, and contain either all the realizations of a word that the listener has encountered (Goldinger, 1998; Johnson, 1997) or consist of a prototypical version of the sound form of a word that represents a blend of all the tokens (Klatt, 1979). The representations of direct-mapping models do not abstract away from information that is encountered in the speech signal and are therefore

highly detailed. Any acoustic difference between the sound form of a word and the sound form of another word therefore constitutes a potential source of information that can affect lexical activation. However, because the level of phonetic detail of lexical representations is, from a theoretical perspective, unconstrained, specifying the dimensions along which similarity computations between the speech signal and lexical representations take place is far from trivial. Perhaps largely owing to this, direct-mapping models have never been fully computationally implemented, which has constrained their role in spoken-word recognition research.

PHONETIC DETAIL IN SPEECH: PROSODY

Although the sound form of a word can be described in an abstract way as a sequence of phonemes, a phonemic (or featural) representation of the speech signal abstracts away from a large amount of systematic, linguistically-determined speech variation. One major source of speech variation is prosody, which is an abstract structure that determines the relative salience and grouping of speech sounds (see Beckman, 1996, and Shattuck-Hufnagel & Turk, 1996 for reviews). The prosodic structure of an utterance consists of a hierarchy of prosodic constituents of different sizes, with lower prosodic constituents (e.g., syllables) being embedded in larger constituents (e.g., words). This structure is manifested in the speech signal by fine-grained yet systematic phonetic variation. The acoustic realization of a speech sound is, for instance, strongly affected by its prosodic position. A given speech sound tends to be of longer duration at the edge of higher prosodic constituents than at the edge of lower prosodic constituents (Ladd & Campbell, 1991; Wightman, Shattuck-Hufnagel, Ostendorf & Price, 1992). Because prosodic constituents equal to or higher than the word are aligned with word boundaries, the lengthening of speech sounds constitutes a cue to the location of a word boundary. However, the usefulness of such cues for spoken-word recognition has often been viewed as marginal. This is because acoustic cues to word boundaries, such as lengthening, are not reliably present in the speech signal. Furthermore, even when such cues are present, they are often small and variable. Nevertheless, when acoustic cues to word boundaries are available in the speech signal, such information could potentially be used by listeners to assist the recognition process. The main goal of the research presented in this thesis was to examine the effects of prosodically-conditioned speech variation on lexical processing. If such information could be shown to affect the recognition of spoken words, this would have important consequences for existing theories and models of spoken-word recognition. In particular, such a finding would demonstrate that the representations involved in

the recognition of spoken words cannot be purely segmental in nature. This would challenge most existing models of spoken-word recognition and, more generally, reveal important constraints on the information that is relevant for lexical processing.

THE EYE-TRACKING PARADIGM

In a seminal and groundbreaking study, Cooper (1974) examined the processing of spoken language in the context of a visual environment. He presented participants with pictures in a visual display and a concurrent spoken story. Cooper found that participants spontaneously fixated relevant pictures in the visual display in response to unfolding referring expressions in the speech stream. For example, they made an eye movement to a picture of a lion upon hearing the word *lion*. Interestingly, fixations to the pictures in the visual display were very closely time locked to relevant information in the story. Participants often fixated a picture that was associated with a spoken word before they had heard the entire word. This suggests that eye movements may reflect the ongoing interpretation of the speech signal and that Cooper's methodology can be applied to the study of the online interpretation and processing of spoken language.

After a gap of about two decades, Tanenhaus, Spivey-Knowlton, Eberhard, and Sedivy (1995) rediscovered the eye-tracking paradigm and applied it to the study of auditory sentence processing. They used a task in which participants' eye movements were recorded while they carried out spoken instructions to move real objects that were arranged on a table in front of the participant. The underlying hypothesis is that as the spoken instruction is heard and processed, gaze direction reflects a participant's ongoing interpretation of the instruction. This version of the paradigm has been used successfully to study a wide range of topics in sentence processing, such as the time course of reference resolution (Eberhard, Spivey-Knowlton, Sedivy, & Tanenhaus, 1995; Sedivy, Tanenhaus, Chambers, & Carlson, 1999), syntactic ambiguity resolution (Tanenhaus et al., 1995; Spivey, Tanenhaus, Eberhard, & Sedivy, 2002) and the use of referential domains (Chambers et al., 2002).

Building on the work of Cooper (1974) and Tanenhaus et al. (1995), Allopenna, Magnuson, and Tanenhaus (1998) extended the use of the eye-tracking paradigm by applying it to the study of spoken-word recognition. They presented participants with a visual display on a computer screen that consisted of four pictures and four geometrical shapes (see Figure 1-1). The task of the participant was to pick up and move one of the objects in accordance with spoken instructions (e.g., "Pick up the beaker. Now put it next to the square."). On most trials, the names of the four objects in the visual

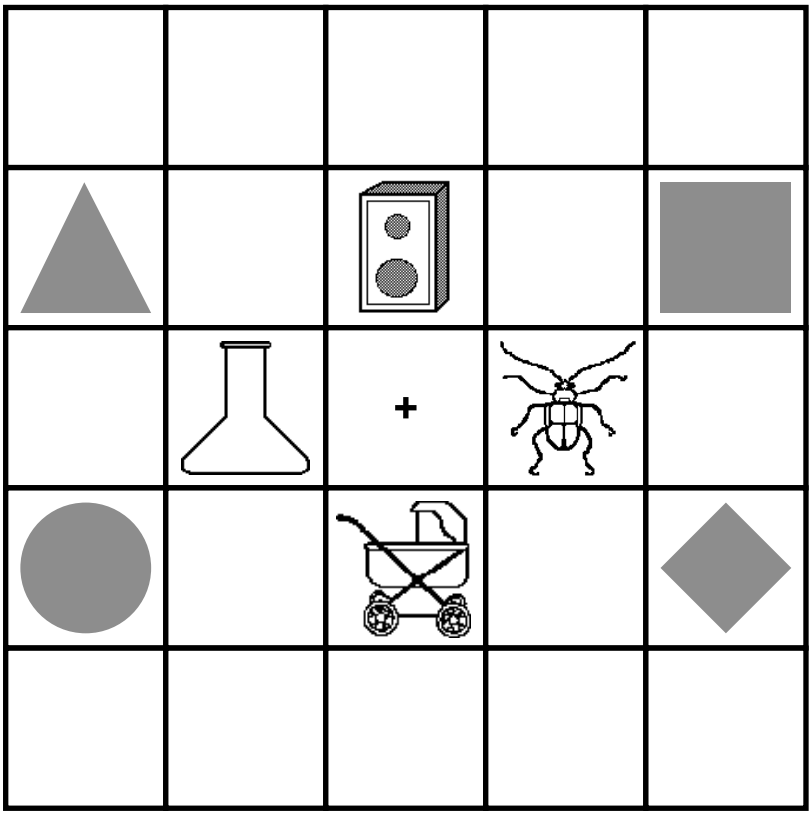

Figure 1-1. An example visual display from Allopenna et al. (1998).

display were phonologically unrelated. However, on experimental trials, the name of one or two of the objects in the visual display was phonologically similar to the name of the referent (e.g., the cohort competitor *beetle*, or the rhyme competitor *speaker*). As the name of the referent unfolds, the speech signal is temporarily consistent with the name of the cohort competitor. If fixations to the objects in the visual display reflect the lexical activation of the names of those objects, one would therefore expect that, upon hearing the name of the referent, participants would be more likely to fixate the picture of the cohort competitor than to fixate either of the phonologically unrelated distractor pictures.

Figure 1-2 presents data from Experiment 1 in the Allopenna et al. (1998) study. Fixation proportions, averaged across participants, are plotted during a time window of a second, starting at the onset of the target word. During this time interval, participants were more likely to fixate the picture associated with the referent, cohort or rhyme than the distractor picture. This strongly suggests that fixations to pictures in the visual display reflect the lexical activation of the names of those pictures. Fixations to the referent and cohort picture started to diverge from fixations to the distractor picture around 200 ms after the onset of the target word, suggesting that eye movements were affected by the processing of the speech input from as early as 200

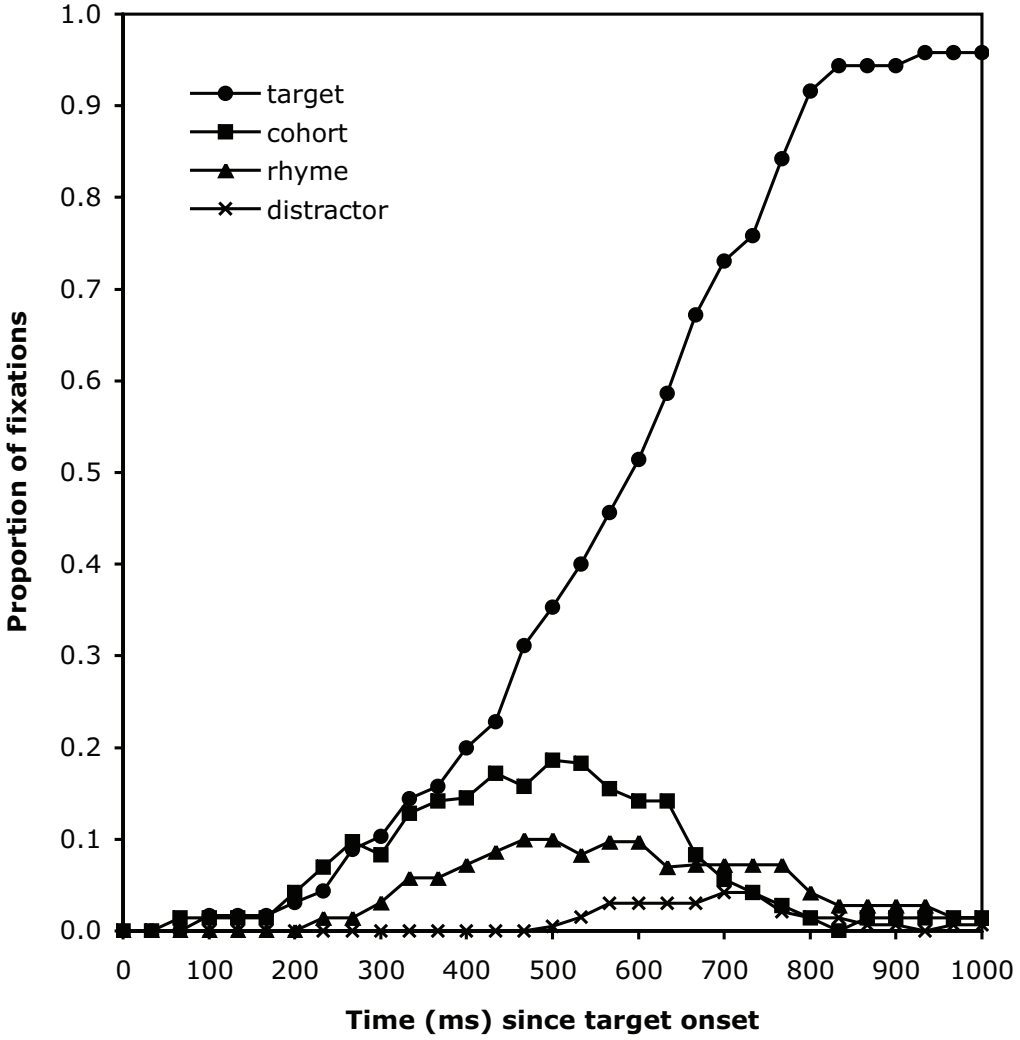

Figure 1-2. Probability of fixating each of the pictures in a visual display over time in Experiment 1 of Allopenna et al. (1998).

ms after the onset of the target word. Taking into account the time it takes to initiate an eye movement, which is estimated to be on average 200 ms (Hallett, 1986), it appears that fixation probabilities reflected changes in lexical activation from the onset of the referent. Further support for the close time locking of input-driven fixations to information in the speech signal was that 200 to 300 ms after phonetic information in the speech signal disambiguated between the referent and the cohort competitor, fixations to the referent started to exceed fixations to that competitor. A further finding of the Allopenna et al. (1998) study was that it provided clear evidence for the activation of rhyme competitors (e.g., *speaker* when the referent is *beaker*), even though such competitors do not overlap with the initial sounds of the referent. Fixations to the rhyme competitor diverged from fixations to the distractor around 300 ms after the onset of the referent.

Allopenna et al. (1998) also conducted simulations with TRACE (McClelland & Elman, 1986) in order to test whether a computational model of spoken-word recognition would predict the observed time course and probabilities of fixations. The input to the model consisted of the name of the referent and lexical activations were computed for the names of each of the objects in the visual display. These activations were subsequently converted to fixation probabilities using the Luce (1959) choice rule. The fixation probabilities that were predicted on the basis of the TRACE simulations provided an exceptionally close fit to the actual fixation probabilities that were observed in the eye-tracking experiment. This seminal finding, that observed fixation probabilities to potential referents in the context of concurrently presented spoken language closely fit predictions of lexical activations derived from TRACE, suggests that the eye-tracking paradigm is a useful tool to study spoken-word recognition.

An important concern about the eye-tracking paradigm used by Allopenna et al. (1998), however, is that their task may encourage participants to develop and make use of strategies. For instance, because a visual display contains only a small set of pictures, participants may adopt a strategy of naming the pictures prior to hearing the instruction sentence, which could potentially allow them to bypass normal lexical processing by evaluating the name of the referent in the context of only the names of the pictures in the visual display. This issue was addressed in two studies that examined whether fixations to referents are sensitive to lexical properties of *non-displayed* items (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Magnuson, Tanenhaus, & Aslin, submitted). Dahan et al. presented participants with a display consisting of a referent (e.g., a net) and three phonologically unrelated distractors. They found that fixations to the referent *net* were delayed when the referent's initial sounds /nɛ/ were replaced by the sequence /nɛ/ from the word *neck* (such that coarticulatory information in the vowel of the referent was consistent with the competitor *neck*, which, importantly, was not present in the visual display), but not when the referent's initial sounds were replaced by the sequence /nɛ/ from the nonword *nep*. An even more compelling demonstration that the eye-tracking paradigm is sensitive to properties of all words in the lexicon concerns a study by Magnuson, Tanenhaus and Aslin (submitted). They used a design in which each visual display consisted of a referent and three phonologically unrelated distractors, and showed that the identification of the referent was affected by the number of words in the lexicon that started with the same initial sounds as the referent. When the initial sounds of the referent were consistent with the onset of many other words, and competition from those other words was therefore expected to interfere strongly with recognition of the referent, fixations to the referent increased more slowly than when the initial sounds of the referent were

consistent with the onset of few other words. This demonstrates that eye movements were influenced by the activation of non-displayed lexical candidates, thus validating the use of the eye-tracking paradigm as a tool to study spoken-word recognition. Fixations to pictures in the visual display do not simply reflect task-specific strategies (e.g., the evaluation of the speech signal in the context of the names of the pictures in the display). Rather, eye movements in the eye-tracking paradigm are sensitive to properties of the normal language-processing system, including patterns of lexical activation across the entire lexicon.

A growing body of research has thus demonstrated the value of the eye-tracking paradigm as a tool for studying the recognition of spoken words in continuous speech. Eye movements provide a continuous measure of lexical activation that is very closely time locked to the speech input, thus providing information about the interpretation of the input at fine-grained temporal resolution, as a spoken word unfolds. Previous research has shown that the paradigm is sensitive to subphonemic information in the speech signal (subcategorical mismatches in Dahan et al., 2001; within-category phonetic variation in McMurray, Tanenhaus, & Aslin, 2002). Furthermore, eye movements are sensitive to patterns of lexical activation across the entire lexicon, capturing subtle and transient effects of activated competitors (Dahan et al., 2001; Magnuson, Tanenhaus, & Aslin, submitted). This renders the paradigm an ideal tool to study the lexical activation and competition process during spoken-word recognition.

SUBPHONEMIC VARIATION

The experiments reported in this thesis examined if and how prosodically-conditioned, subphonemic speech variation affects spoken-word recognition. Previous research has demonstrated that lexical activation is sensitive to subphonemic variation. One line of research has examined the influence of mismatching acoustic-phonetic information on lexical activation (e.g., Dahan et al., 2001; Marslen-Wilson & Warren, 1994; McQueen, Norris, & Cutler, 1999). These studies have demonstrated that lexical processing can be affected by mismatching coarticulatory information. For instance, listeners are slower in activating the lexical representation of the word *net* when coarticulatory information in the vowel /ɛ/ is inconsistent with the upcoming segment /t/ (for instance, when the initial sounds /nɛ/ of the word *net* originate from the word *neck*, as in Dahan et al., 2001). Other studies have demonstrated that subphonemic variation that occurs in natural speech can have an impact on lexical actvation. Andruski, Blumstein, and Burton (1994) found that a token of

the word *peach* activated the word *peach* more strongly than an edited version of the same token whose voice-onset-time (VOT) had been reduced (for related and similar results see van Alphen & McQueen, in press; McMurray, Tanenhaus, & Aslin, 2002; Utman, Blumstein, & Burton, 2000).

The aforementioned studies have established that subphonemic variation can affect lexical activation. However, all of these studies used speech that was artificially manipulated. The impact of naturally-occurring (i.e., unedited) subphonemic variation on lexical activation has only recently become a topic of investigation. Gow (2002), for instance, presented listeners with an assimilated version of the word *right* in the context of the phrase *right berries*. In this context, the word *right* is usually realized closely resembling the form /raɪp/, when the place of articulation of its final consonant /t/ assimilates to the place of assimilation of the following phoneme /b/. Gow showed that listeners, when presented with an assimilated version of the word *right*, activated the lexical representation of the word *right* more strongly than that of the word *ripe*. The phonetic realization of the assimilated form /raɪp/ thus preserved some characteristics of its underlying form /raɪt/, and listeners' lexical interpretation of the word was sensitive to these naturally-occurring subphonemic details.

The studies on the effect of subphonemic variation on lexical activation discussed so far considered subphonemic variation that is phonemically contrastive, i.e. information that affects the degree of phonemic support for lexical candidates. For instance, in the Andruski et al. (1994) study, variation in VOT of the phoneme /p/ or /b/ translates to support for the interpretation of the segment as unvoiced (i.e., /p/, when VOT is relatively long) or voiced (i.e., /b/, when VOT is relatively short). The main contribution of these studies is that they demonstrated that subphonemic information in the speech signal can have an impact on lexical activation. The results are, however, not inconsistent with models of spoken-word recognition that rely on phonemic representations. Such models can account for the findings of the aforementioned studies, provided that their phonemic representations can be activated in a graded fashion (for a detailed discussion of this issue, see McQueen, Dahan, & Cutler, 2003).

The research presented in this thesis examined the influence on lexical activation of subphonemic speech variation that is not phonemically contrastive. If such information could be shown to have an impact on lexical processing, this would provide important information about the nature of the representations involved in recognizing spoken words. In particular, such a finding would challenge existing models that rely on phonemic representations. For instance, when the acoustic realization of a particular word across different prosodically-defined positions is phonemically identical, but

phonetically different, models of spoken-word recognition that rely on a strictly phonemic encoding of the speech signal predict that the position of the word should not have a strong impact on its identification. Demonstrating that prosodically-conditioned speech variation affects lexical processing would therefore have important theoretical implications for theories and models of spoken-word recognition.

STRUCTURE OF THE THESIS

This thesis examines the influence of speech variation conditioned by prosodic structure on the recognition of spoken words. Most current models of spoken-word recognition assume that such information should not have a systematic impact on lexical processing. Chapter 2 investigated whether listeners can use acoustic information associated with prosodic structure to discriminate onset-embedded words from their longer competitors. Although phonemically identical, the spoken sequence /hæm/ tends to be of longer duration when it corresponds to the short word *ham* (because it is followed by a prosodic-word boundary) than when it corresponds to the onset of a longer word, for example, *hamster*. Three eye-tracking studies examined whether listeners can use subphonemic acoustic cues in the speech signal to discriminate embedded words from their longer competitors. Chapter 3 focussed on the influence of prosodically-conditioned variation in the realization of one and the same word on lexical processing and describes two eye-tracking experiments that contrasted the recognition of words across different, prosodically-defined positions: in utterance-medial and utterance-final position. This study thus compared the processing of the same word across different, prosodically-defined positions, while the study in Chapter 2 asked whether the word-recognition system is sensitive to prosodically-conditioned acoustic differences between different words occuring in the same position in an utterance. Chapter 4 consists of a summary of the findings of this thesis.

REFERENCES

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

Alphen, P. M. van, & McQueen, J. M. (in press). The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance.*

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

Beckman, M. E. (1996) The parsing of prosody. *Language and Cognitive Processes, 11*, 17—67.

Chambers, C. G., Tanenhaus, M. K., Eberhard, K. M., Filip, H., & Carlson, G. N. (2002). Circumscribing referential domains during real-time language comprehension. *Journal of Memory and Language, 47*, 30—49.

Cooper, R. M. (1974). The control of eye fixation by the meaning of spoken language. A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84—107.

Cutler, A., & Norris, D. (1988). The role of strong syllables in segmentation for lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 14*, 113—121.

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.

Eberhard, K. M., Spivey-Knowlton, M. J., Sedivy, J. C., & Tanenhaus, M. K. (1995). Eye movements as a window into real-time spoken language comprehension in natural contexts. *Journal of Psycholinguistic Research, 24*, 409—436.

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

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

Gow, D. W., 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.

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.

Johnson, K. (1997). 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.

Klatt, D. H. (1979). Speech perception: A model of acoustic-phonetic analysis and lexical access. *Journal of Phonetics, 7*, 279—312.

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, France.

Luce, R. D. (1959). *Individual choice behavior*. New York: Wiley.

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

Magnuson, J. S., Tanenhaus, M. K., & Aslin, R. N. (submitted). Time and similarity of spoken words.

Marslen-Wilson, W. D. (1973). Linguistic structure and speech shadowing at very short latencies. *Nature, 244*, 522—523.

Marslen-Wilson, W. D. (1975). Sentence perception as an interactive parallel process. *Science, 189*, 226—228.

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.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29—63.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33—B42.

McQueen, J. M. (1998). Segmentation of continuous speech using phonotactics. *Journal of Memory and Language, 39*, 21—46.

McQueen, J. M., Dahan, D. & Cutler, A. (2003). Continuity and gradedness in speech processing. In A. S. Meyer & N. O. Schiller (Eds.), *Phonetics and phonology in*

language comprehension and production: Differences and similarities (pp. 39—78). Berlin: Mouton de Gruyter.

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.

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

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191—243.

Sedivy, J. C., Tanenhaus, M. K., Chambers, C. G., & Carlson, G. N. (1999). Achieving incremental semantic interpretation through contextual representation. *Cognition, 71*, 109—147.

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

Shillcock, R. (1990). Lexical hypotheses in continuous speech. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 24—49). Cambridge, MA: MIT Press.

Spivey, M. J., Tanenhaus, M. K., Eberhard, K. E., & Sedivy, J. C. (2002). Eye movements and spoken language comprehension: Effects of visual context on syntactic ambiguity resolution. *Cognitive Psychology, 45*, 447—481.

Tabossi, P., Burani, C., & Scott, D. (1995). Word identification in fluent speech. *Journal of Memory and Language, 34*, 440—467.

Tanenhaus, M. K., Spivey-Knowlton, M. J., Eberhard, K. M., & Sedivy, J. C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science, 268*, 1632—1634.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62*, 1297—1311.

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.

THE ROLE OF PROSODIC BOUNDARIES IN THE RESOLUTION OF LEXICAL EMBEDDING IN SPEECH COMPREHENSION

CHAPTER 2

Anne Pier Salverda, Delphine Dahan, and James M. McQueen (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51—89.

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.

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. 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 instan-

tiated 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 informa-

tion 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) *F*0 differences. The mean *F*0 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 naïve to the purpose of the study confirmed the presence of durational and *F*0 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" (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 *F*0 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 *F*0 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; Oller, 1973; Martin, 1970, for English; Nooteboom & Doodeman, 1980; Cambier-Langeveld, 2000, 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, and 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, *F*0, 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 2.1). A Dutch speaker, naïve 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]

[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

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 2.2 and 2.3). 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 [stuk-gesneden/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 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.

---

or the Beat Addition rule, resulting in lengthening the first syllable or pausing between the two syllables (see Liberman & Prince, 1977; Selkirk, 1984).

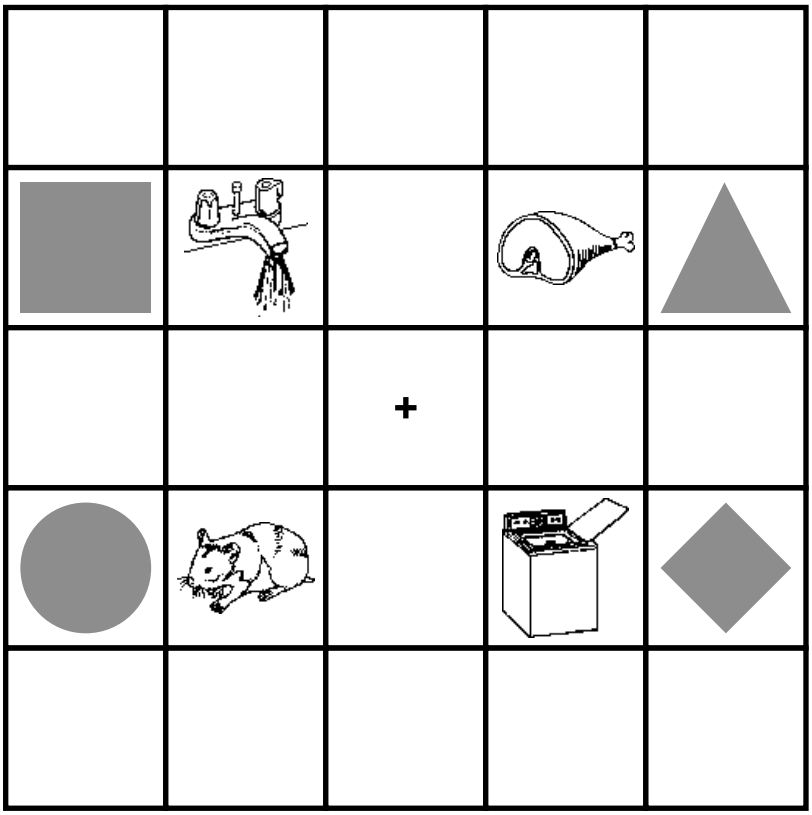

Figure 2-1. Example of a visual display. The geometrical shapes were green.

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 Figure 2-1) 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, 2001b), as well as with the lexical frequency associated with the picture's

name (Dahan, Magnuson, & Tanenhaus, 2001a). 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.

EXPERIMENT 2.1

Experiment 2.1 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 2.1 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 2.1A) or in an unstressed-syllable context (Experiment 2.1B).

METHOD

Participants

Sixty native speakers of Dutch, students at the University of Nijmegen, participated in the experiment (30 in Experiment 2.1A, 30 in Experiment 2.1B).

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 2-1). 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 the exception of the vowel /ʉ/, which was substituted for the reduced vowel /ə/ in 4 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

Table 2-1. 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 2.1 (the underlined portion of each sentence was used to create the cross-spliced versions).

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

latter, the syllable was unstressed in 23 out of the 28 items; for the remaining 5 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 software. 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 2.1A, comparing carrier-word and monosyllabic-word stressed-context conditions, and Experiment 2.1B, 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. Moreo-

[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.

ver, 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.

Acoustic analyses

The duration of the sequences, as well as the mean fundamental frequency (*F*0) 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 2.1A) 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 2.1B) 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 ms and 15 ms) than in the Davis et al. (2002) study (48 ms). Measures of the mean *F*0 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).

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 Figure 2-1) 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 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 2.1A or 2.1B), 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.

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/second), and acceleration (8000 degrees/second^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.

RESULTS

The goal of Experiment 2.1 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

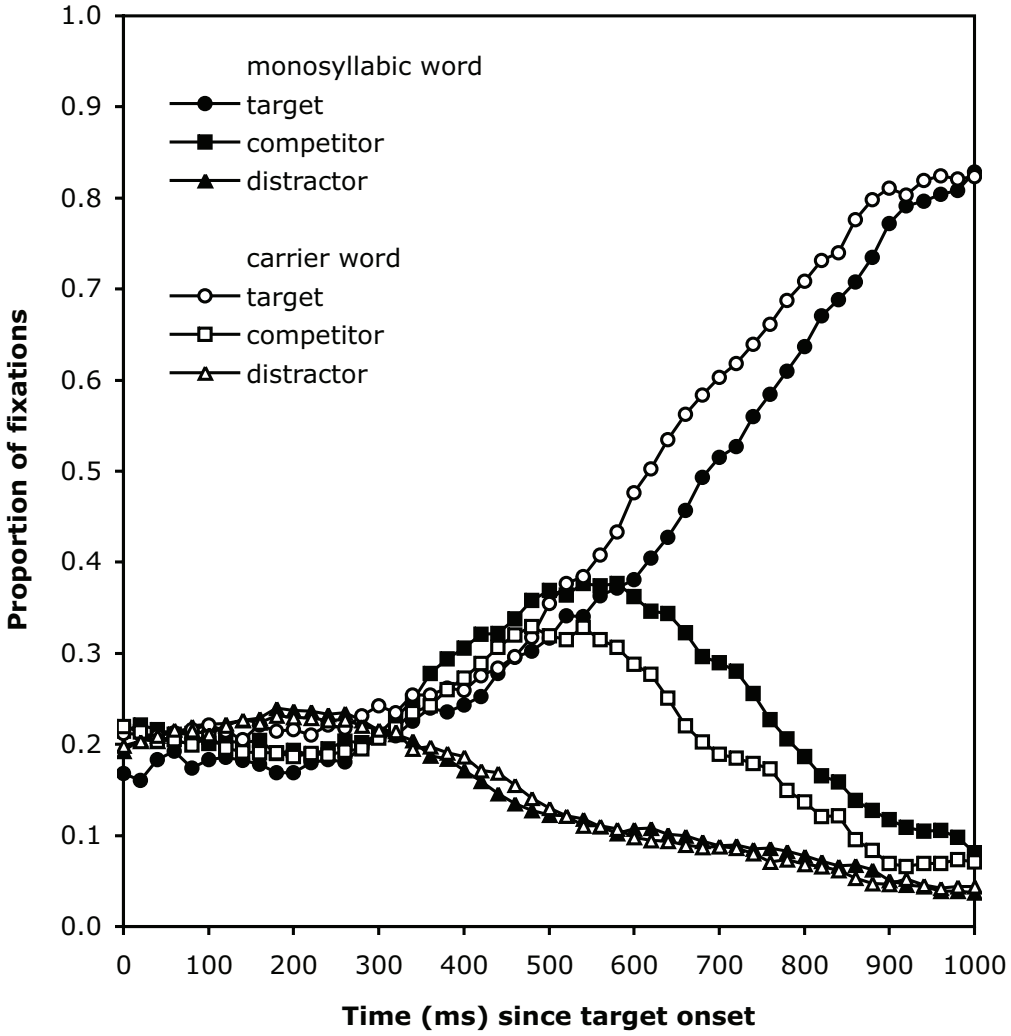

Figure 2-2. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 2.1A (carrier-word vs. monosyllabic-word stressed-context condition).

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 2.1A), or from the same token of the carrier word and from a monosyllabic word followed by an unstressed syllable (Experiment 2.1B).

Experiment 2.1A

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 dur-

ing 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).

Figure 2-2 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 the target word. As is apparent from the graph, the proportions of fixations to any picture on the display were equivalent at 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 measure. 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 ($F_1(1,29) = 11.6$, $p < .005$; $F_2(1,27) = 5.5$, $p < .05$).

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 Figure 2-2, 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 ($F_1(1,29) = 9.6$, $p < .005$; $F_2(1,27) = 4.9$, $p < .05$), and a main effect of Window ($F_1(1,29) = 23.2$, $p < .001$; $F_2(1,27) = 10.1$, $p < .005$), but the interaction did not reach significance ($F_1(1,29) = 1.9$, $p > .10$; $F_2(1,27) = 3.1$, $p > .05$). 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 Figure 2-2 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 Figure 2-2, 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 ms 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 et al., 2001a, 2001b). 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 the General Discussion.

Experiment 2.1B

Experiment 2.1B was identical to Experiment 2.1A 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. Figure 2-3 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 2.1A. 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 2.1A.

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 ($F_1(1,29) = 5.9$, $p < .05$; $F_2(1,27) = 1.5$, $p > .10$), 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 ($F_1(1,29) = 65.7$, $p < .001$; $F_2(1,27) = 19.1$, $p < .001$), an effect of Condition significant only by participants ($F_1(1,29) = 5.2$, $p < .05$; $F_2(1,27) = 1.4$, $p > .10$), and no interaction (F_1 and $F_2 < 1$).

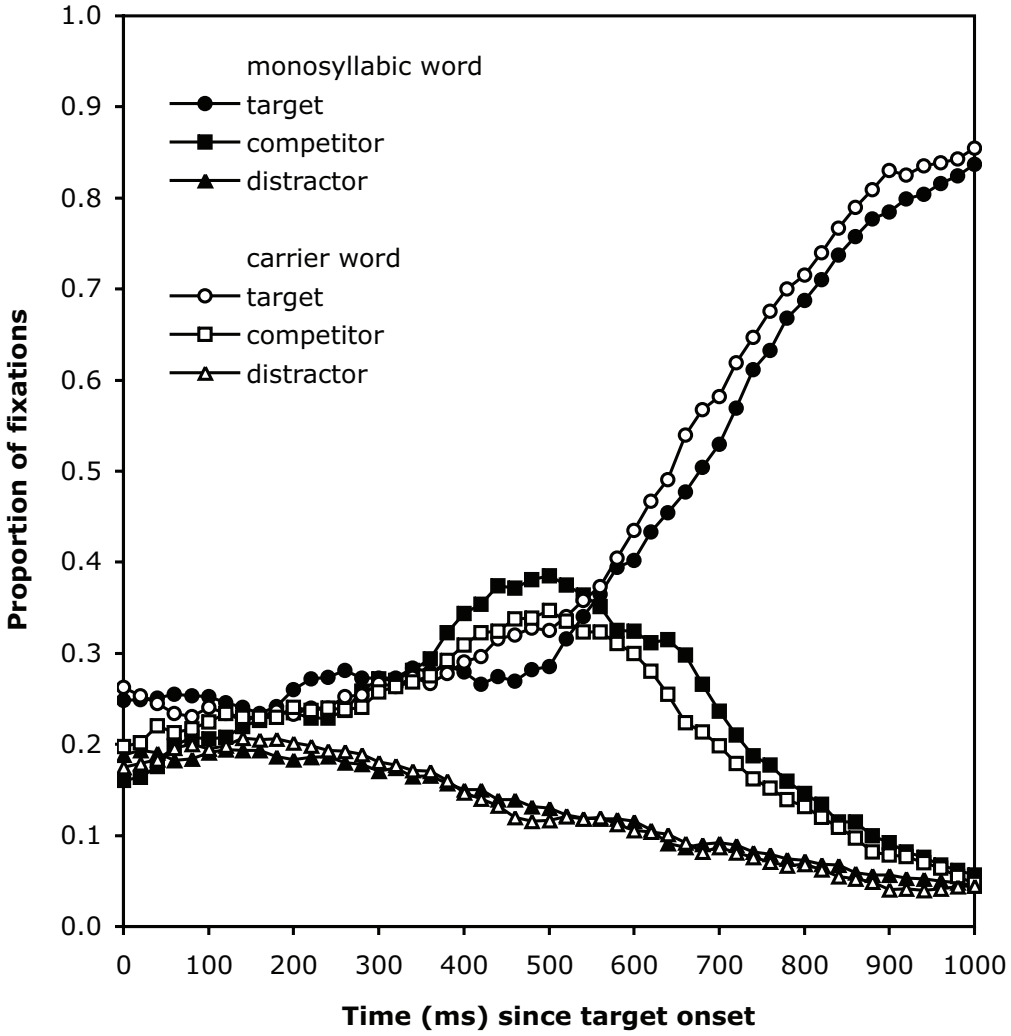

Figure 2-3. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 2.1B (carrier-word vs. monosyllabic-word unstressed-context condition).

In order to compare the pattern of results from Experiments 2.1A and 2.1B, 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 ($F_1(1,58) = 17.4$, $p < .001$; $F_2(1,27) = 4.8$, $p < .05$), 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, the inter-item variability across items observed in Experiment 2.1B but not in Experiment 2.1A (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.

DISCUSSION

Experiment 2.1 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 2.1A) and by an unstressed syllable (Experiment 2.1B).

Experiment 2.1A 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 2.1B 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 Figures 2-2 and 2-3: The difference in competitor fixations between the monosyllabic- and carrier-word conditions was smaller in Experiment 2.1B than in Experiment 2.1A. 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 2.1B, with the same statistical power as Experiment 2.1A 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 the introduction, 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 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 be-

[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.

[4] An alternative explanation for the difference in lexical disambiguation between Experiments 2.1A and 2.1B 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 4 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 2.2 and 2.3, where the duration of the sequence, rather than the context in which it was originally produced, biased its lexical interpretation.

tween 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 2.1A: $r(26) = .61$, $p < .001$; for Experiment 2.1B: $r(26) = .54$, $p < .005$; for both experiments: $r(54) = .59$, $p < .001$). 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 *F*0 differences between monosyllabic- and carrier-word stimuli (from naïve and non-naïve 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 naïve speakers but not by the non-naïve 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 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 2.1A'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 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 2.1A 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 2.1A, 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 2.2) or opposite to Experiment 2.1A's pattern (Experiment 2.3). 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

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.

EXPERIMENT 2.2

Experiment 2.2 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.

METHOD

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 2.1A or 2.1B.

Materials and Procedure

Our stimuli were selected from the same source as the stimuli used in Experiment 2.1A. 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

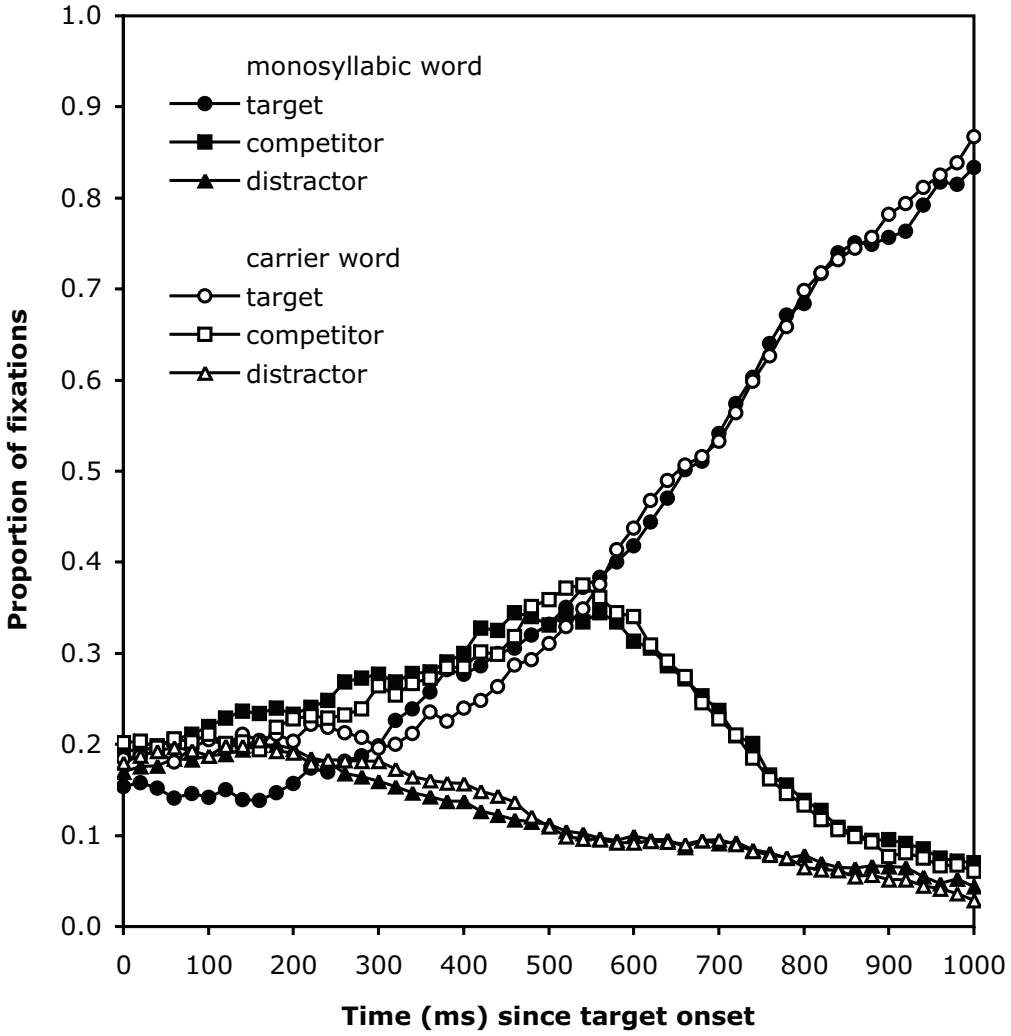

Figure 2-4. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 2.2.

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 2.1A (245 ms). Measures of the mean *F*0 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 2.1. Design, procedure, and coding were the same as in Experiment 2.1.

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 (3 trials, 0.4% of the data). Figure 2-4 presents the proportion of fixations over time to the target 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 2.1. 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 2.1A and Experiment 2.2. 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 ($F_1(1,58) = 4.2$, $p < .05$; $F_2(1,27) = 3.4$, $p = .08$), no main effect of Experiment, and, importantly, a significant interaction between Condition and Experiment ($F_1(1,58) = 4.0$, $p < .05$; $F_2(1,27) = 4.2$, $p = .05$).

In Experiment 2.2, 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 2.1A'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 condition 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 2.1A, 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 2.2, no bias in lexical interpretation is observed.

EXPERIMENT 2.3

Experiment 2.3 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 2.3 was thus reversed from the pattern present in Experiment 2.1A'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.

METHOD

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.

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 7 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 7 items were included in the experiment, but excluded from all analyses. There was a negligible difference in the mean *F*0 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 2.1 and 2.2.

RESULTS AND DISCUSSION

On a few trials, participants erroneously moved the competitor picture rather than the target picture (3 out of 630 trials, 0.5% of the data). These trials were excluded from the analyses. Figure 2-5 presents the proportion of fixations to the target picture, to the competitor 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 ($F_1(1,29) = 2.2$, $p > .10$; $F_2(1,20) = 1.5$, $p > .10$). A two-way (Condition × Time Window [300-550 ms vs. 550-900 ms]) ANOVA revealed no main effect of Condition ($F_1(1,29) = 1.2$, $p > .10$; $F_2(1,20) < 1$), a main effect of Window ($F_1(1,29) = 22.8$, $p < .001$; $F_2(1,20) = 16.5$, $p < .005$) and, crucially, a significant interaction ($F_1(1,29) = 4.6$, $p < .05$; $F_2(1,20) = 6.7$, $p < .05$). 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%; $F_1(1,29) = 8.5$, $p < .01$; $F_2(1,20) = 7.4$, $p < .05$). 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

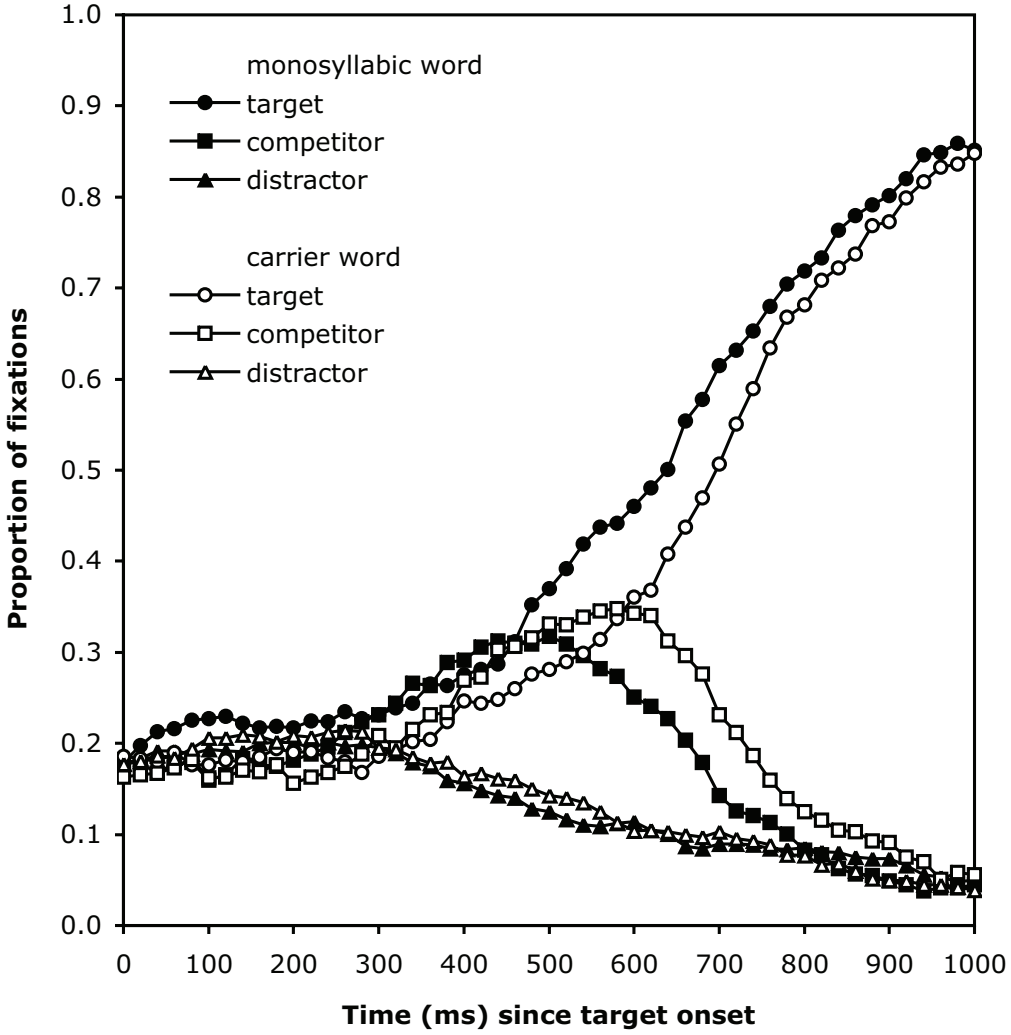

Figure 2-5. Proportion of fixations over time for the target, competitor, and averaged distractors, for the monosyllabic-word condition and the carrier-word condition in Experiment 2.3.

550-900 ms interval between these two conditions ($r(19) = .54$, $p < .01$; this correlation was also significant for the 300-900 ms time interval, $r(19) = .72$, $p < .001$).

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 2.1A and Experiment 2.3, after excluding from the Experiment 2.1A data the seven items that were excluded from the Experiment 2.3 analysis. 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 ($F_1(1,58) = 8.8$, $p < .005$; $F_2(1,20) = 11.7$, $p < .005$), a non-significant effect of Condition, and a significant interaction ($F_1(1,58) = 13.5$, $p < .005$; $F_2(1,20) = 11.9$, $p < .005$).

Experiment 2.3 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 2.1A on the basis of their duration, we were able to make the fixation pattern observed in Experiment 2.1A reverse. This confirms the importance of sequence duration in modulating the lexical interpretation of ambiguous sequences.

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 2.1 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 2.1A, where the monosyllabic word had been followed by a stressed syllable in its original recording, it was smaller and not fully significant in Experiment 2.1B, where the monosyllabic word had been followed by an unstressed syllable. Nevertheless, the statistically non-significant interaction between Experiments 2.1A and 2.1B 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 2.1's results suggest that the disambiguation of an embedded se-

quence is subject to variability that the lexical origin of the embedded sequence could not account for.

Experiment 2.2 replicated Experiment 2.1A 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 2.3, 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 2.1'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 originated 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 the Introduction), which has shown that prosodic boundaries influence the duration of pre-boundary 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 2.1 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 2.1A and the inter-item variability observed in Experiment 2.1B.

In the Introduction, 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 2.1A and 2.1B 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 syn-

tactic 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 2.2) and segmental lengthening does not always coincide with a word boundary, presumably caused by other prosodic phenomena such as pitch accents (Experiment 2.3). 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 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

---

[5] Measurements of the formant frequencies *F*1 and *F*2 on the sequences' vowels in the monosyllabic-word and carrier-word conditions in Experiment 2.1 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 2.1, analyses of the *F*1 and *F*2 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 *F*1/*F*2 values for each of the 9 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 9 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 ($p > .05$). The same analyses performed on the formant frequencies of the sequences' vowels in Experiments 2.2 and 2.3 showed differences in vowel space that were non-significant and, importantly, inconsistent with the tendency found in Experiment 2.1 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.

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 pointed out in the discussion of Experiment 2.1A, 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 et al., 2001a, 2001b). 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 interpreta-

tion 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 et al., 2001b; 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.

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. (2001a). 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. (2001b). 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, France.

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 (Ph.D. 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.

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 (miscroscope)
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
Ze dacht dat die HAMSTER verdwenen was.	ˈhɑm.stəʀ
Ze dacht dat die HAM stukgesneden was.	ˈhɑm ˈstʉk.χə.sneː.də
Ze 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 dəˈzɛlv.də
Ik geloof dat die KANDELAAR er niet meer is.	ˈkɑn.dəˈlaːʀ
Ik geloof dat die KAN dubbel zo veel kostte.	ˈkɑn ˈdʉ.bəl
Ik geloof dat die KAN dezelfde kleur heeft.	ˈkɑn dəˈzɛlv.də
Hij had die KIJKER meegenomen.	ˈkɛi.kəʀ
Hij had die KEI kundig ingepakt.	ˈkɛi ˈkʉn.dəχ
Hij had die KEI kunstzinnig beschilderd.	ˈkɛi kʉnstˈsɪ.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 səˈkyːʀ
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 kʉlˈtyːʀ.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.ʊə
Zij dacht dat die SCHIL dezelfde vorm zou hebben.	ˈsχɪl də.ˈzɛlv.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 χə.ˈʀʉst
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 kʉnst.ˈmaː.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 χə.ˈlʉ.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

THE INFLUENCE OF PROSODICALLY-CONDITIONED SPEECH VARIATION ON THE EVALUATION OF LEXICAL CANDIDATES IN SPOKEN-WORD RECOGNITION

CHAPTER 3

The research presented in this chapter was carried out while the author was a visiting graduate student at the University of Rochester. This work was done in collaboration with Delphine Dahan (MPI and University of Pennsylvania), and Michael Tanenhaus, Mikhail Masharov, Katherine Crosswhite, and Joyce McDonough (University of Rochester).

INTRODUCTION

Spoken-word recognition involves the simultaneous and partial activation of candidate words in response to the unfolding speech signal (e.g., Marslen-Wilson, 1987; Zwitserlood, 1989). For example, the initial sounds of the word *cap*, /kæ/, are potentially consistent with multiple candidate words, including the words *cat*, *captain* and *candy*. As a spoken word unfolds, the candidate words that most closely resemble the speech signal most strongly interfere with the recognition of the spoken word (e.g., Luce, 1986; Marslen-Wilson, 1993). The identification of a spoken word is therefore constrained, over time, by the extent to which its sound form resembles that of other words.

The idea that lexical processing involves the parallel activation and consideration of multiple candidates that are (at least partly) compatible with the speech input was first embodied in the influential Cohort model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson, 1987). In this model, the recognition of a word was viewed as the on-line process of distinguishing it from other words that are consistent with the unfolding spoken input. Luce and colleagues (Luce, 1986; Luce, Pisoni, & Goldinger, 1990; Luce & Pisoni, 1998) altered this view in an important way by demonstrating that the candidates that are acoustically similar to the spoken word interfere with its recognition. The more frequent and/or numerous those candidates, the more interference arises. The activation and competition process is therefore determined by the degree of similarity between the speech input and candidates that are considered for recognition, as a word unfolds. The present study examined whether variation in the re-

alization of a spoken word has an impact on the recognition process by affecting the evaluation of lexical candidates.

One central characteristic of speech that is often pointed out is its variability. No two tokens of a particular spoken word sound exactly alike, even if they are produced by the same speaker in the same context. How does variation in the realization of a spoken word influence its recognition? Most current psycholinguistic theories and models of spoken-word recognition assume, implicitly or explicitly, that only information in the speech signal that distinguishes a word from other words is relevant to the recognition process. Variation in the realization of a particular spoken word that does not affect its similarity to other words is the kind of information that the recognition process must abstract away from. That is, in order to attribute different realizations of a spoken word to the same lexical entry, the information associated with contextual variation should be neutralized during lexical access. Because phonemes, or more precisely, the set of distinctive phonetic features that compose a particular phoneme, have traditionally been viewed as the minimal lexically contrastive units, variations that do not affect the phonemic (or featural) interpretation of the input are viewed as irrelevant to the process of lexical processing. An exception to this view is formed by exemplar-based theories (e.g., Goldinger, 1998; Johnson, 1997). These theories assume that all acoustic details of every word token, including lexically irrelevant characteristics such as the speaker's voice, are maintained in memory. Lexical processing thus involves computing the similarity between the memory traces of all exemplars and the acoustic signal. Perhaps because of the complexity in specifying the metrics involved in the similarity computation, these models are still relatively marginal in spoken-word recognition research.

The traditional phonemic/featural view outlined above has had important consequences for the modeling of spoken-word recognition. First, phonemes and/or features are viewed as mediating the mapping of the acoustic signal onto the lexicon. Second, the degree to which candidate words are activated upon hearing a particular spoken input is assumed to be invariant across contexts so long as the phonemic or featural interpretation of the input remains the same. For example, when listeners hear a version of the spoken word *cap* in any utterance context, candidate words that overlap at onset (i.e., cohort competitors), such as *captain* and *cat*, will strongly compete for recognition with *cap*. Furthermore, the word *captain* will be a stronger competitor than the word *cat* because *captain* overlaps with *cap* by a greater stretch of the input—computed in terms of distinctive features or phonemes—than *cat* does.

We argue here that this view needs to be revised. The present study provides evidence that the relative degree of activation of candidate words varies as a function of the prosodic context in which a given to-be-recognized word occurs.

It is well established that variations in the phonetic characteristics of phonemes affect lexical processing. For example, artificially varying the Voice Onset Time (VOT) of an English word's initial voiceless stop consonant affects the activation of the word's lexical representation: As the VOT moves away from the prototypical value of a voiceless stop consonant and toward the prototypical value of a voiced stop consonant, a decrease in the activation of the intended word (e.g., pear) is observed, in tandem with an increase in the activation of its voiced counterpart, if it corresponds to an existing word (e.g., bear) (Andruski, Blumstein, & Burton, 1994; McMurray, Tanenhaus, & Aslin, 2002; Utman, Blumstein, & Burton, 2000; see also van Alphen and McQueen, in press, for related findings in Dutch). The impact of naturally occurring fine-grained variation in the realization of phonemes on lexical activation has also been demonstrated by Gow (2002). An optional phonological assimilation rule in English constrains a coronal phoneme, such as /t/, to adopt the place of articulation of the stop consonant that follows—provided that both consonants belong to the same phrase. For example, in the phrase *right berries*, the place of articulation of the final consonant of the word *right* becomes bilabial, the place of articulation of the following phoneme /b/. According to the assimilation rule, /t/ should be realized closely resembling its bilabial counterpart /p/, leading to a lexical ambiguity between *right* and *ripe* or even to a misinterpretation of *right* as *ripe*. However, Gow showed that listeners, when presented with the assimilated version of the word *right* in the context of *berries*, activate the lexical representation of *right* more strongly than that of *ripe*. Gow thus demonstrated that assimilation was incomplete. This means that the phonetic realization of the assimilated final segment was a poor exemplar of the segment /p/ and preserved some characteristics of its underlying form, the coronal /t/, and that lexical processing reflected these phonetic details.

The studies just reviewed have documented how lexical activation is modulated by acoustic information that may affect the degree of support for a particular phonemic interpretation of the spoken input. These findings are compatible with the idea that the mapping of the speech input onto representations of lexical form is mediated, and therefore constrained, by a phonemic encoding of the spoken input (but only if the phonemic encoding is probabilistic; see McQueen, Dahan, and Cutler, 2003). However, some more recent studies have shown that lexical activation can be modulated by variations in the signal that do not affect the phonemic interpretation of the speech signal. Davis, Marslen-Wilson, and Gaskell (2002), for English, and Salverda,

Dahan, and McQueen (2003), for Dutch, showed that the interpretation of a lexically ambiguous word fragment (e.g., /hɑm/) as corresponding in Dutch either to a monosyllabic word (e.g., *ham*, id.) or to the first syllable of a longer word (e.g., *hamster*, id.) is influenced by the duration of the fragment. Longer fragments tend to be assigned a monosyllabic interpretation more than shorter fragments do. This bias reflects the durational distribution of ambiguous fragments: In both studies, measurements of naturally produced utterance tokens revealed longer averaged fragment durations when the token was produced as a monosyllabic word than when it was produced as the initial syllable of a polysyllabic word. The consistency across these studies, and across the speakers within the Davis et al. study, suggests that this durational difference is a robust feature of both English and Dutch.

Thus, lexical processing appears to make use of systematic variations that an encoding of the signal in terms of phonemes alone would not capture (see Spinelli, McQueen, & Cutler, 2003, for findings on the phonological phenomenon *liaison* in French that provide other evidence of this claim). Building on this, the present study addresses the broader impact of these findings on the recognition of spoken words by showing that the competition environment of a spoken word is not fixed, but dynamically established as a function of the fine-grained, prosodically-conditioned details of the spoken input. A word that is a strong competitor of a spoken word in one prosodic position may not be a strong competitor of the same word in another prosodic position. This contrasts with the view that a given word is always associated with the same set of competitors—and that the degree to which those competitors are considered for recognition is not affected by variation in the word's realization as associated with its prosodic position. The two findings of the present study—the inadequacy of a purely phonemic analysis and the dynamic nature of the lexical competition process—force us to reconsider the processes and representations mediating the mapping of the acoustic signal onto the lexicon.

A well-studied source of systematic variation in the realization of an utterance's segments is the utterance's prosodic structure. This is an abstract structure that determines the relative saliency and grouping of speech units (for reviews, see Beckman, 1996; Shattuck-Hufnagel & Turk, 1996). According to theories of prosodic organization (Selkirk, 1984; Beckman & Pierrehumbert, 1986; Nespor & Vogel, 1986), this structure consists of a hierarchy of constituents of different sizes. Lower-level constituents (e.g., syllables) are embedded into larger constituents at an immediately higher level, up to the highest level of prosodic constituency (usually referred to as the Utterance). The prosodic structure of an utterance also manifests itself in *pitch accents*, which indicate the prominence status of words in the utterance. The presence

of a pitch accent on a word affects the acoustic and phonetic characteristics of its segments (including its duration), which in turn can affect lexical processing. For instance, an accented word is processed more rapidly than its deaccented counterpart, despite the fact that the duration of the accented word, and thus the time that is needed to access the spoken input, is greater (e.g., Cutler, 1976; Cutler & Foss, 1977). It has been proposed that the presence of a prominence enhances the salience of the word's phonetic features by reducing coarticulation and by reducing the overlap in the acoustic cues specifying distinct phonemic categories (Bard, 1990; Cho, 2002; Cole & Jakimik, 1980; see also Kuhl et al., 1997, for such a demonstration on infant-directed speech, which is characterized by large pitch excursions). This would in turn reduce the activation of spurious candidates and their interference with the recognition of the spoken word.

The prosodic structure of an utterance is also apparent in the phonetic details of the segments located at constituent edges. For example, a well-established correlate of prosodic structure is the lengthening of speech segments in preboundary position (Edwards, Beckman, & Fletcher, 1991; Klatt, 1976; Oller, 1973). The amount of lengthening of preboundary segments has been shown to vary depending on the size of the prosodic constituent. For instance, final lengthening is stronger for segments that occur at the edge of an utterance than for segments at the edge of a lower constituent such as the prosodic word (Ladd & Campbell, 1991; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). Furthermore, prosodic boundaries most strongly affect the realization of speech segments that appear in their immediate vicinity. For instance, utterance-final lengthening primarily affects the rhyme of the final syllable of the utterance (Cambier-Langeveld, 2000, for Dutch; Klatt, 1976; Oller, 1973; Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992, for English) and, within the syllable, it affects the coda more than the nucleus (Campbell & Isard, 1991). Thus, the domain over which preboundary lengthening applies does not necessarily correspond to a lexical word. A monosyllabic word (e.g., *cap*) is almost entirely affected by preboundary lengthening, but a longer word (e.g., *captain*) is mostly affected on its final syllable and virtually not on its initial syllable.

The fact that the domain of preboundary lengthening is not lexically defined has interesting consequences for the characteristics of lexically ambiguous fragments, such as /kæp/, which can occur as a monosyllabic word or embedded at the onset of longer words (e.g., *captain*). Such a sequence tends to be affected by preboundary lengthening differently depending on whether it corresponds to a monosyllabic word or to the first syllable of a polysyllabic word. In the former case, the sequence immediately precedes a prosodic-constituent edge and therefore undergoes lengthening,

while in the latter case, the sequence is further away from a prosodic-constituent edge and is hardly affected by preboundary lengthening. The present study aimed to demonstrate that listeners make use of the differential impact of preboundary lengthening on words of different (syllabic) length in the course of interpreting lexically ambiguous fragments such as /kæp/. This finding would support the view that lexical processing, rather than abstracting away from subphonemic yet systematic variations, makes use of these variations in computing the goodness of fit between the acoustic signal and sound-form lexical representations.

In this study, we manipulated the amount of preboundary lengthening that a spoken word underwent by varying its position within the utterance's prosodic structure. The spoken word appeared in utterance-final position, as in "Now click on the cap", or in utterance-medial position, as in "Put the cap next to the square". In utterance-final position, the segments of the syllable at the edge of the prosodic constituent were expected to be substantially lengthened. In utterance-medial position, however, the critical word was aligned with a lower prosodic-constituent edge (i.e., a Prosodic-Word boundary), and was therefore predicted to undergo substantially less preboundary lengthening. By varying the position of the critical word, we induced naturally-occurring durational variations of the critical word, and as noted above, these variations were expected to affect monosyllabic and polysyllabic words differently. We therefore examined the effect of prosodically-conditioned durational variations on the recognition of monosyllabic and polysyllabic words. This effect was evaluated by considering both the time course of the recognition of monosyllabic or polysyllabic spoken words and the degree to which monosyllabic or polysyllabic competitor words were considered for recognition. We asked whether the processing of an ambiguous sequence (e.g., /kæp/) that has undergone preboundary lengthening results in favoring monosyllabic interpretations (e.g., *cap*) over polysyllabic interpretations (e.g., *captain*). Furthermore, we asked whether the processing of an ambiguous sequence that does not display evidence of preboundary lengthening (e.g., the initial syllable of *captain* in utterance-final position) results in disfavoring monosyllabic interpretations (e.g., *cap*).

EXPERIMENT 3.1

Experiment 3.1 examined how the processing of a spoken word—the speed with which it is identified and the degree to which spurious onset-overlapping competitors are considered in the course of processing the spoken word—varies as a function of prosodically-conditioned variations in its acoustic realization. Monosyllabic and poly-

syllabic words, such as *cap* and *captain*, were produced in utterance-medial and utterance-final positions. The degree of activation of monosyllabic and polysyllabic competitor words when listeners heard the monosyllabic or polysyllabic spoken word in each of the utterance positions was assessed by collecting and analyzing listeners' eye gaze to pictures as they followed spoken instructions to manipulate (using a computer mouse) one of four pictured objects displayed on a computer screen. The referent picture's name was a monosyllabic or polysyllabic word, occurring either in utterance-medial position (e.g., "Put the cap/captain next to the square") or in utterance-final position (e.g., "Now click on the cap/captain"). The display consisted of the picture of the referent object, of a competitor object whose name overlapped at onset with the referent's name and was either polysyllabic or monosyllabic, and of two objects with unrelated names. Eye movements to displayed objects are taken to reflect listeners' on-going interpretation of the spoken input, based on the assumption that people direct their attention and gaze toward objects in order to guide a mouse movement toward the referent object. This paradigm has proved to provide a fine-grained measure of lexical processing and competition over time (Allopenna, Magnuson, & Tanenhaus, 1998; Dahan, Magnuson, & Tanenhaus, 2001a; Dahan, Magnuson, Tanenhaus, & Hogan, 2001b; Salverda et al., 2003; see Tanenhaus, Magnuson, Dahan, & Chambers, 2000, and references therein). The degree to which the process of identifying the referent's name leads to the temporary activation of a competitor's name is estimated by the probability that listeners shift their gaze toward the competitor picture. Furthermore, the speed of recognition of the referent's name can be estimated by the timing of listeners' eye movements to the referent picture prior to clicking on it.

Experiment 3.1 considered three related questions. First, we tested the hypothesis that the preboundary lengthening that affects a monosyllabic word such as *cap* in utterance-final position results in decreased support for a polysyllabic word candidate, such as *captain*. This effect, we argue, results from the association between preboundary lengthening and the syllabic structure of the upcoming word: Preboundary lengthening occurring at the right edge of a high prosodic constituent affects the duration of the entire monosyllabic word, whereas it affects the first sounds of polysyllabic words only minimally. Thus, a long speech fragment /kæp/ would be a poorer match to a polysyllabic competitor, like *captain*, than a shorter speech fragment /kæp/, despite their equivalent phonemic match. The predicted decreased activation of a polysyllabic competitor word upon hearing a monosyllabic word in utterance-final, as opposed to utterance-medial position should translate into a lower probability of making an eye movement toward the competitor picture. Modulation of competitor

fixations as a function of the position of the monosyllabic spoken word may also affect fixations to the referent picture. If the competitor in the display represents a weaker competitor, listeners' interpretations may converge toward the target picture faster.

The *decreased* activation of polysyllabic competitors as a result of preboundary lengthening of a monosyllabic spoken word (i.e., when the monosyllabic spoken word was in utterance-final, as opposed to utterance-medial position) was contrasted with a predicted *increased* activation of monosyllabic competitors. When the syllabic structures and overall length of the spoken word and the competitor match, we argue, an increase of the duration of the portion of the spoken word that is consistent with the competitor should translate into an increase in the strength of evidence supporting this competitor. Thus, according to our second hypothesis, the activation of a monosyllabic competitor like *cat*, and therefore listeners' probability of generating an eye movement toward the competitor picture, should be greater when the monosyllabic spoken word *cap* was lengthened (i.e., in utterance-final position) than when it was not (i.e., in utterance-medial position). Greater consideration of the competitor picture in the utterance-final condition should in turn affect the speed with which listeners converge toward the target picture. Observing *increased* activation of monosyllabic competitors and concurrent *decreased* activation of polysyllabic competitors when the monosyllabic spoken word undergoes preboundary lengthening would provide compelling evidence that lexical activation reflects prosodically-conditioned durational variations present in the speech input.

Finally, we tested whether the degree of activation of a monosyllabic competitor word, such as *cap*, upon hearing a polysyllabic referent word, such as *captain*, was affected by the position of the polysyllabic spoken word within the utterance. As mentioned above, the initial sounds of a polysyllabic word are hardly affected by preboundary lengthening. Consequently, in utterance-final position, the fragment /kæp/ that constitutes the initial sounds of *captain* is not affected as much by preboundary lengthening as it would be if the monosyllabic word *cap*, rather than *captain*, had been produced. We asked whether the relatively short duration of the initial sounds of a polysyllabic word in utterance-final position is information that can be used to disfavor the interpretation of the unfolding word as a monosyllabic word. This condition differs from its counterpart (i.e., when the spoken word is monosyllabic and the competitor is polysyllabic) in an interesting way. The lack of preboundary lengthening is not in itself incompatible with the presence of a monosyllabic word. Indeed, in the utterance "Click on the cap there", the fragment /kæp/ would probably not show substantial lengthening, and in fact should be of a comparable duration to the same frag-

ment in the utterance "Click on the captain". Nevertheless, if listeners are led to anticipate that the ambiguous fragment belongs to the last word of the utterance (perhaps because all utterances that start with "Click on the__" in the experimental context end with the referent's name), the absence of preboundary lengthening on the ambiguous fragment may constrain lexical processing in a similar way as its presence. Accordingly, the activation of a monosyllabic competitor would decrease when the polysyllabic spoken word was in utterance-final vs. utterance-medial position.

Our hypotheses on the effect of the spoken word's position in the utterance on competitor activation focus on how preboundary lengthening may constrain the set of competitors as a function of their syllabic structure (the first and third hypotheses), or increase the activation of competitors matched in syllabic structure by increasing the duration of the ambiguous fragment (the second hypothesis). However, varying the position of the spoken word affected more than just the duration of its last segments. In particular, words in utterance-medial position were deaccented in order to avoid the production of a major prosodic boundary that often follows words that receive a pitch accent. Words in final position, however, were accented. This difference in accent pattern associated with the position of the referent in the instruction sentence is expected to have an impact on the recognition of spoken words, with words in utterance-final position being identified faster than words in utterance-medial position. This applies equally to monosyllabic and polysyllabic referents because a pitch accent primarily affects the realization of a word's syllable bearing primary stress, and the pitch accent associated with words in utterance-final position was therefore expected to have a similar effect on the realization of the initial sounds of monosyllabic and polysyllabic referents. Fixations to target pictures were expected to reflect the difference in accent pattern associated with the position of the referent.

METHOD

Participants

Twenty-four students of the University of Rochester took part in the experiment. They were all native speakers of American English and were paid a small amount for their participation.

Materials

Forty word pairs were selected from the CELEX lexical database (Baayen, Piepenbrock, & Gulikers, 1995). All words were picturable nouns. Twenty of these pairs consisted of two monosyllabic nouns that only differed in the place of articulation of

their final consonant (e.g., *cap* and *cat*, *comb* and *cone*). The remaining twenty pairs consisted of a polysyllabic noun (e.g., *candy*) and a monosyllabic noun that matched the initial sounds of the polysyllabic word (e.g., *can*). Each of the 40 word pairs was paired with two phonologically unrelated distractor words to form four-word sets, with the additional constraint that each set comprised two monosyllabic words and two polysyllabic words (e.g., candy, can, helmet, saw). The 40 stimulus sets are listed in Appendix A. Pictures representing the four words from each stimulus set were selected from various picture databases.

In addition to the 40 experimental stimulus sets, 50 filler sets were constructed. Each set consisted of a target word and three distractor words that were phonologically unrelated to the target word. Within each filler set, two words were monosyllabic and two polysyllabic. To discourage participants from developing expectations, based on the experimental sets, that pictures with similar names in the display were likely targets, 40 of the 50 filler sets had two similar-sounding distractor words (20 of them monosyllabic words differing only on their last consonant, [e.g., *lock* and *log*] and 20 with one word embedded in the other [e.g., *spy* and *spider*]). Within each filler set, one word, never one of the two similar-sounding words, was selected to play the role of referent. Within the 90 trials of the experiment, half of the referent words were monosyllabic and half polysyllabic. Pictures for the filler trials were selected from the same databases as those used for the experimental trials.

For each of the 40 experimental stimulus sets, two instruction sentences were constructed for each of the two words of a pair, yielding a total of four sentences per stimulus set (see Table 3-1). These sentences varied which item of a word pair was the target word, as well as the target word's position within the utterance. The target word could appear in utterance-medial position (e.g., "Put the cap next to the square") or in utterance-final position (e.g., "Now click on the cap"). The same sentence frames were used to construct instruction sentences for the 50 filler sets. In half of these sentences, the target word occurred in utterance-medial position, and, for the other half of the sentences, in utterance-final position.

All sentences were recorded on digital audiotape in a quiet room, using a head-mounted microphone. The female speaker, who was a trained phonetician, read the sentences in a randomized order. In order to minimize the realization of a prosodic break after the target word in utterance-medial position—which would lead to pre-boundary lengthening on the target word and compromise the effectiveness of our position manipulation—the speaker was instructed to produce each sentence as one intonational phrase. This resulted in a pitch accent on the stressed syllable of the last

Table 3-1. Example of a set of instruction sentences for each condition in Experiment 3.1.

Monosyllabic referents with monosyllabic competitors (e.g., cap/cat)

	Instruction sentence	Competitor
Utterance-medial condition	Put the cap next to the square	cat
	Put the cat next to the square	cap
Utterance-final condition	Now click on the cap	cat
	Now click on the cat	cap

Monosyllabic referents with polysyllabic competitors (e.g., can/candy)

	Instruction sentence	Competitor
Utterance-medial condition	Put the can next to the square	candy
Utterance-final condition	Now click on the can	candy

Polysyllabic referents with monosyllabic competitors (e.g., candy/can)

	Instruction sentence	Competitor
Utterance-medial condition	Put the candy next to the square	can
Utterance-final condition	Now click on the candy	can

word of the sentence, thus with no pitch accent on the target words in utterance-medial position.

All sentences were then digitized and labeled using a speech editor. Durational measurements were made on polysyllabic and monosyllabic target words in utterance-medial and utterance-final positions. Table 3-2 presents the average duration of the onset, nucleus, and coda (when the monosyllabic word had them) of the monosyllabic word and of the corresponding segments in the polysyllabic words, in utterance-medial and utterance-final position. Monosyllabic words were markedly longer in utterance-final position, when they immediately preceded the utterance edge, than in utterance-medial position (320 ms vs. 262 ms, a 22% increase, collapsing over the two types of monosyllabic referents). By contrast, the increase in duration for fragments corresponding to the same segments in polysyllabic words was noticeably more modest. On average, the fragment was 207 ms in utterance-medial position and 214 ms in utterance-final position, an increase of only 3%. The measurements for the duration of the onset, nucleus and coda of the items reported in Table 3-2 confirm the expectation based on the phonetics literature that final lengthening would most strongly affect the realization of segments immediately preceding the utterance boundary.

Table 3-2. Mean duration (in ms) of the segments of the target word in the utterance-medial (e.g., "Put the cap next to the square") and utterance-final (e.g., "Now click on the cap") condition of Experiment 3.1. Numbers of observations within each cell are indicated in parentheses.

	Monosyllabic referent (e.g., *cap*); monosyllabic competitor			
	Utterance-medial	Utterance-final	Difference	Lengthening
Onset	75 (38)	70 (38)	-5	-7%
Nucleus	124 (38)	157 (38)	34	27%
Coda	73 (38)	118 (34)	45	62%
Total	272 (38)	333 (38)	61	22%

	Monosyllabic referent (e.g., *can*); polysyllabic competitor			
	Utterance-medial	Utterance-final	Difference	Lengthening
Onset	69 (19)	65 (19)	-4	-5%
Nucleus	132 (20)	171 (20)	39	29%
Coda	68 (14)	98 (13)	29	43%
Total	245 (20)	296 (20)	51	21%

	First syllable of polysyllabic referent (e.g., *[can]dy*); monosyllabic competitor			
	Utterance-medial	Utterance-final	Difference	Lengthening
Onset	61 (19)	60 (19)	-1	-1%
Nucleus	113 (20)	118 (20)	5	5%
Coda	52 (14)	57 (14)	5	9%
Total	207 (20)	214 (20)	7	3%

Note. There are different numbers of observations across cells because in the monosyllabic referent, polysyllabic competitor condition, 7 items did not have either an onset or a coda (e.g., *ant*, *antler*; *tie*, *timer*) and because the final sound of 5 monosyllabic referents ending in /t/ was not released in utterance-final position (e.g., *cat*).

Design

Experiment 3.1 consisted of a total of 90 trials (40 experimental trials and 50 filler trials). An experimental trial consisted of the presentation of the pictures associated with a stimulus set along with one of the four instruction sentences. The 40 experimental trials consisted of 20 trials with monosyllabic target words and monosyllabic competitors, 10 trials with monosyllabic target words and polysyllabic competitors and 10 trials with polysyllabic target words and monosyllabic competitors. Four lists

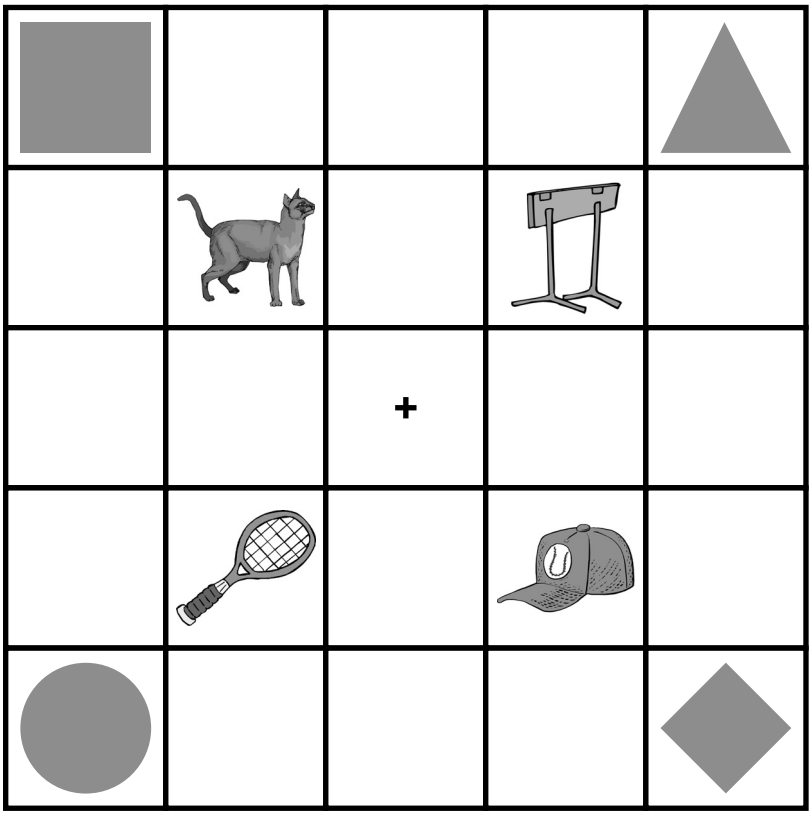

Figure 3-1. Example of a visual display in Experiment 3.1.

were constructed by varying which of the four sentences that were recorded for every experimental stimulus set was presented along with the visual display. Within each list, the target word occurred in utterance-medial position in half of the experimental trials and in utterance-final position in the other half. The order of trials was pseudo-randomized, with three filler trials at the beginning to familiarize participants with the procedure.

Procedure

Participants were seated at a comfortable distance from a computer screen. Eye movements were monitored using a head-mounted Applied Sciences Laboratories E5000 eye tracker. A small scene camera aligned with the participant's line of sight provided a continuous recording of the visual scene. Prior to the experiment, the eye-tracking system was calibrated, allowing software to superimpose a participant's point-of-gaze on a HI-8 videotape recording of the scene provided by the scene camera, at a rate of 30 frames per second. The spoken sentences were presented to participants through headphones and simultaneously recorded on the videotape.

Two computers controlled the presentation of stimuli during the experiment. One computer was used to present spoken sentences over headphones, while the other computer presented the accompanying visual display. The experimenter triggered the presentation of these events by pressing the computers' spacebars. The structure of each trial was as follows. First, a 5×5 grid appeared on the computer screen, with a fixation cross in the center, shortly followed by a pre-recorded instruction to fixate the central cross ("Look at the cross"). This allowed assessment of the accuracy of the eye-tracker calibration. Then, the experimenter triggered the appearance on one computer of a visual display, which was composed of four pictures and four geometric shapes (see Figure 3-1). After a short delay, the critical instruction was presented (e.g., "Put the cap next to the square") through the experimenter pressing the second computer's spacebar.

Coding procedure

An editing VCR with frame-by-frame controls was used to examine the videotape recording of each participant, and hence to establish which of the pictures in the visual display were fixated as the target sentence unfolded. Fixations were coded for each frame on the videotape, starting at the onset of the target word up to and including the time frame when the saccade to the target object that preceded the initiation of a mouse movement to the target object was initiated. The crosshair superimposed on the scene camera's recording of the visual scene was used to establish, for each frame, whether the participant fixated the target picture, the competitor picture, one of the two distractor pictures, or another location on the computer screen. When a saccade was being performed, and thus no gaze data was available, the corresponding time frames were assigned to the target location of the saccade.

RESULTS

Two pairs of items from the monosyllabic referent, monosyllabic competitor condition (*mouth-mouse*, *sheep-sheet*) were discarded because all participants who were presented with the referents *mouth* and *sheet* in final condition erroneously moved the competitor picture (i.e., a mouse or sheep, respectively). In this condition, a total of 47 other trials (corresponding to 10.3% of the data) were discarded because participants moved or clicked on the picture of the monosyllabic competitor without correcting their choice (33 trials in medial condition and 14 trials in final condition). A few trials were discarded because of technical failure or track loss (10 out of a total of 696

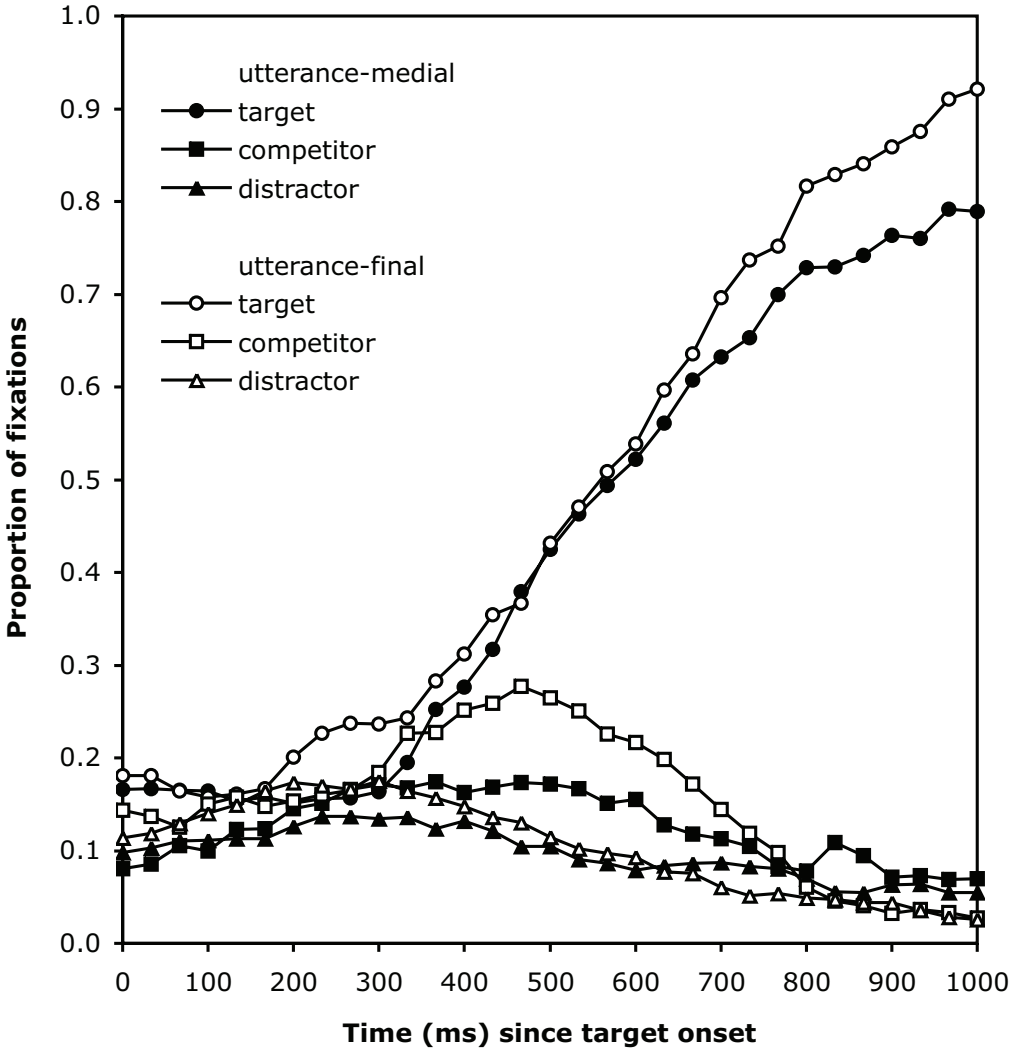

Figure 3-2. Proportion of fixations over time to the target, the competitor, and the averaged distractors, in utterance-medial and utterance-final position, for monosyllabic referents with monosyllabic competitors in Experiment 3.1.

trials; 1.4% of the data) and two trials were discarded because participants failed to carry out the spoken instruction.

Fixation probabilities to each type of picture (the target, competitor, and averaged for the two distractors) were computed for each condition. This was done by adding, for each participant, for each 33-ms time interval starting at target-word onset, the total number of trials on which a particular type of picture was fixated, and dividing this number by the total number of trials on which, during the same time interval, any location on the screen was fixated. Figures 3-2, 3-3 and 3-4 present, for each of the three conditions of Experiment 3.1 (i.e., monosyllabic referents with monosyllabic competitors in Figure 3-2, monosyllabic referents with polysyllabic competitors in Figure 3-3, and polysyllabic referents with monosyllabic competitors in Figure 3-4),

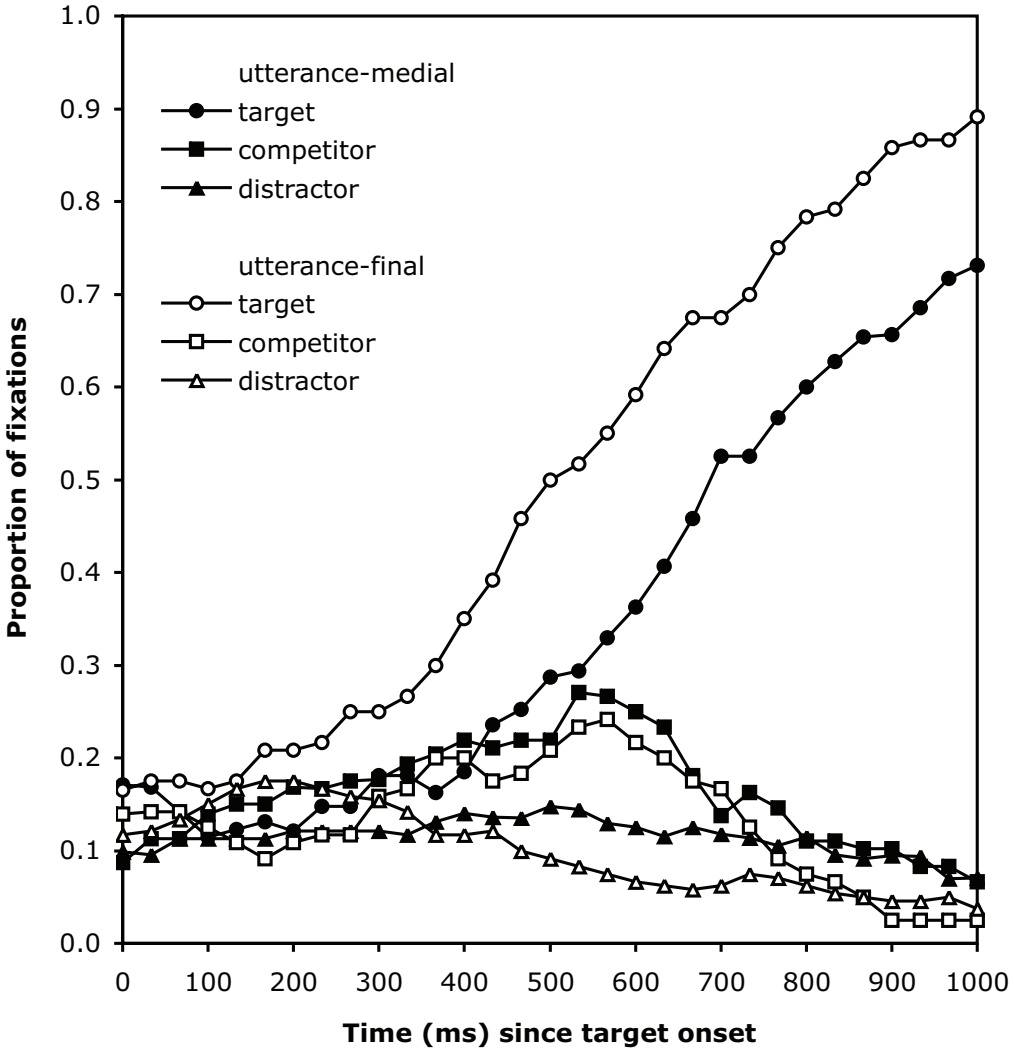

Figure 3-3. Proportion of fixations over time to the target, the competitor, and the averaged distractors, in utterance-medial and utterance-final position, for monosyllabic referents with polysyllabic competitors in Experiment 3.1.

fixation probabilities to the target, the competitor and the averaged distractor, in medial and final position, from 0 to 1000 ms after the onset of the target word.

DESCRIPTIVE SUMMARIES

Monosyllabic referents with monosyllabic competitors

For monosyllabic referents with monosyllabic competitors, around 200 ms after the onset of the target word, fixation proportions to the target picture began to rise in both medial and final condition (see Figure 3-2). Target fixations between conditions were comparable until around 800 ms after the onset of the target word, when fixation proportions to the target picture started to increase more rapidly in final than in medial

position. However, during the time interval over which no difference in fixation proportions to the target as a function of position was observed, the position of the referent strongly affected fixations to the monosyllabic competitor. Fixation proportions to the competitor, in both medial and final position, began to rise around 200 ms after the onset of the target word. Competitor fixations started to diverge from distractor fixations around 400 ms in medial position and around 300 ms in final position, and remained higher than distractor fixations until shortly after 800 ms. In line with our predictions, fixation proportions to the competitor increased faster and reached a higher peak in final than in medial position, indicating that the monosyllabic competitor was considered for recognition more strongly when the monosyllabic referent occurred in final position than when it occurred in medial position. Interestingly, the effect of position on target fixations appears to arise around 800 ms after the onset of the target word, that is, immediately after the effect of position on competitor fixations has ended. This suggests that the absence of an effect of position on target fixations until around 800 ms may result from the fact that monosyllabic competitors competed for recognition with the target words more strongly in final position than in medial position and that this competition acted to suppress the advantage for targets in final position that emerged after 800 ms.

Monosyllabic referents with polysyllabic competitors

For monosyllabic referents with polysyllabic competitors, fixation proportions to the target picture began to rise around 200 ms after the onset of the target word in medial position and slightly earlier in final condition (see Figure 3-3). There was a major effect of position such that, from as early as 200 ms after target-word onset, participants were more likely to fixate the target picture when the referent occurred in final position than when it occurred in medial position. The magnitude of the effect of position on fixation proportions to the target increased steadily, from 200 ms after target-word onset, and was greatest from 500 to 700 ms. Competitor fixations began to rise and started to diverge from distractor fixations shortly before 200 ms after the onset of the target word in medial position, and shortly after 200 ms after target-word onset in final position. Competitor fixation proportions remained higher than distractor fixation proportions until around 800 ms in both medial and final position. Fixation proportions to the competitor started to rise earlier and reached a slightly higher peak in medial than in final position. This effect of the position of the referent on the time spent fixating the competitor (which is the reverse of what was observed in the monosyllabic referent, monosyllabic competitor condition) is in line with our predictions and indicates that the polysyllabic competitor was considered for recognition more

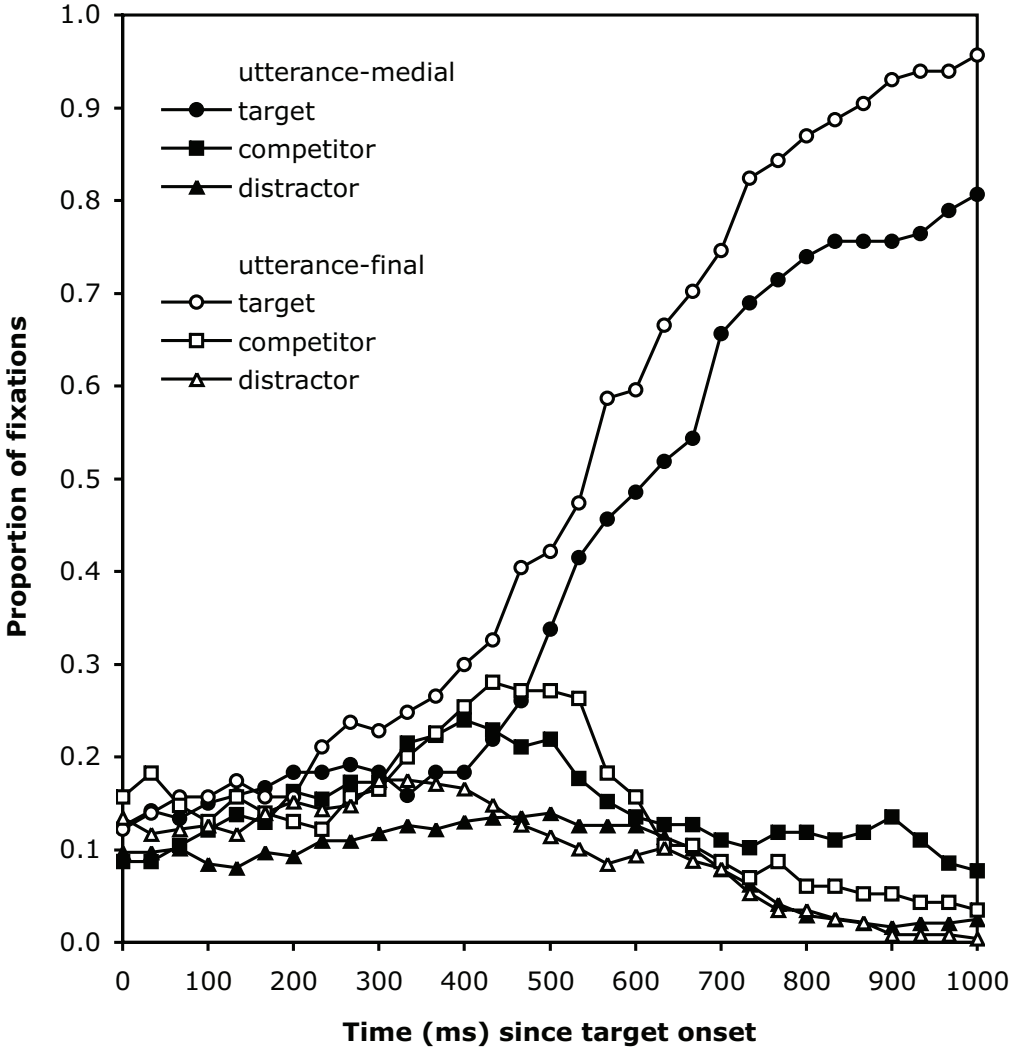

Figure 3-4. Proportion of fixations over time to the target, the competitor, and the averaged distractors, in utterance-medial and utterance-final position, for polysyllabic referents with monosyllabic competitors in Experiment 3.1.

strongly when the monosyllabic referent occurred in medial position than when it occurred in final position.

Polysyllabic referents with monosyllabic competitors

For polysyllabic referents with monosyllabic competitors, around 200 ms after the onset of the target word, fixation proportions to the target started to rise in both medial and final position (see Figure 3-4). There was an effect of position such that target fixations increased more rapidly in final position than in medial position. This effect was apparent as early as shortly after 200 ms after the onset of the target word, and was strongest from around 700 ms. Competitor fixations began to rise shortly after 200 ms after target-word onset in both medial and final position, and started to diverge from distractor fixations around 200 ms in medial position, and shortly before

400 ms in final position. Fixation proportions to the competitor reached a higher peak in final position than in medial position. In both medial and final position, competitor fixations merged with distractor fixations around 600 ms after the onset of the target word. In medial position, competitor fixations diverged again from distractor fixations from shortly after 700 ms after target-word onset until shortly after 1000 ms.

STATISTICAL ANALYSES

To examine the identification of the target word and the activation of the target and competitor as the target word was heard and processed, several types of analyses were performed on the eye-movement data. Because it is estimated that it takes on average 200 ms to program an eye movement (Hallett, 1986), fixations from 200 ms after the onset of the target word were assumed to reflect the lexical activation of the names of the pictures that were represented in the visual display. Therefore, fixations to the pictures that occurred before 200 ms after target-word onset were discarded in all the analyses.

Analyses were performed for target and competitor fixations separately, averaging across participants and items. For some analyses, the time interval over which a measure was computed was unconstrained, taking into account all fixations that occurred from 200 ms after the onset of the target word until the end of the trial (i.e., until the initiation of a mouse movement toward the target picture), thus providing a global measure of lexical activation during an entire trial. However, the drawback of such analyses is that they provide little information about the locus and time course of any effects that are observed. Therefore, other analyses took into account only fixations during a particular time window that itself was motivated by the time course of fixation probabilities over time. These analyses examined the effect of the position of the referent on the time spent fixating the referent (or the competitor) over a time interval during which the proportion of fixations to the competitor (in both medial and final position) exceeded the proportion of fixations to the distractor. It was assumed that during this time interval, the competitor competed for recognition with the referent, and fixations to the referent and competitor picture thus reflected the lexical competition process. For monosyllabic competitors of monosyllabic referents, and for polysyllabic competitors of monosyllabic referents (see Figures 3-2 and 3-3), this time window extended from 200 to 800 ms after the onset of the target word, while it extended from 200 to 600 ms after the onset of the target word for monosyllabic competitors of polysyllabic targets. In addition to these measures of target and competitor fixation probabilities, target fixation latencies were also computed. Target fixation

Table 3-3. Mean latency (in ms) to fixate the target picture in utterance-medial and utterance-final position, for each condition in Experiment 3.1. Standard errors are indicated in parentheses.

Target, Competitor	Example	Utterance-medial	Utterance-final
monosyllabic, monosyllabic	cap, cat	885 (59)	718 (52)
monosyllabic, polysyllabic	can, candy	1042 (76)	708 (33)
polysyllabic, monosyllabic	candy, can	847 (45)	611 (35)

latency was the latency with which a participant fixated the target picture, prior to initiating a mouse movement toward the picture. This measure was assumed to reflect the ease with which the target word was identified. Average target fixation latencies were computed across participants and items, discarding trials for which the latency to fixate the target picture was less than 200 ms.

To summarize, the following types of statistical analyses were performed for both experiments reported in this paper, comparing the processing of a referent in utterance-medial vs. utterance-final position. A target fixation latency analysis was run to examine the identification of the referent. Analyses that considered the time spent fixating the competitor picture examined the degree to which processing of the referent was associated with the consideration for recognition of different types of competitors. These different types of analyses were performed in order to show that they converged on showing the same effects on lexical processing of the position of the referent.

Target identification

Table 3-3 presents the average target fixation latencies for all three conditions of Experiment 3.1. Target words were identified more rapidly in final position than in medial position. This was observed in all three conditions of the experiment. The latency to fixate the target was significantly shorter when the target word occurred in final position than when the target word occurred in medial position (monosyllabic referents with monosyllabic competitors: $t_1(23) = 3.2$, $p < .005$, $t_2(35) = 3.8$, $p < .001$; monosyllabic referents with polysyllabic competitors: $t_1(23) = 4.4$, $p < .001$, $t_2(19) = 3.9$, $p < .001$; polysyllabic referents with monosyllabic competitors, $t_1(23) = 6.0$, $p < .001$, $t_2(19) = 4.2$, $p < .001$).

This finding could simply reflect the fact that accented words (in final position) are processed more rapidly than unaccented words (in medial position). However, the identification of the target word was expected to be further influenced by the degree to which a displayed competitor was considered for recognition, as a function of the

referent's position in the utterance. In particular, a crucial prediction was that the position of a monosyllabic target word would affect the activation of monosyllabic and polysyllabic competitors differently. Monosyllabic competitors of monosyllabic referents were expected to compete for recognition more strongly when the target word occurred in final position than in medial position. Conversely, it was expected that polysyllabic competitors would compete for recognition more strongly when the referent occurred in medial position than when it occurred in final position. According to these predictions, and if the degree to which those competitors were considered for recognition influenced identification of the target word, target fixation latencies in final position should be more strongly reduced, compared to these latencies in medial position, when the competitor was polysyllabic than when the competitor was monosyllabic.

Indeed, target fixation latency for monosyllabic referents in final position was reduced more strongly (compared to target fixation latency in medial position) when the competitor was polysyllabic (on average 334 ms), than when the competitor was monosyllabic (on average 168 ms). A two-way ANOVA on average target fixation latencies for monosyllabic referents was conducted, with the factors Position (medial vs. final) and Type of Competitor (monosyllabic vs. polysyllabic). Type of Competitor was a within-participants factor in the analysis by participants and a between-items factor in the analysis by items. The interaction between Position and Type of Competitor, though numerically large, was not significant ($F_1(1,23) = 4.0$, $p = .06$; $F_2(1,54) = 1.4$, $p = .24$). The latency to fixate the target was thus not significantly affected by whether the displayed competitor was monosyllabic or polysyllabic.

In order to examine whether changes in competitor activation as a function of the position of a monosyllabic referent were associated with a concurrent effect on fixations to the referent, the average time spent fixating monosyllabic referents was computed over a time interval from 200 to 800 ms after the onset of the referent, that is, during the time interval when fixation proportions to competitors were higher than fixation proportions to distractors. During this time interval, the average time spent fixating referents with monosyllabic competitors was 237 ms in medial position and 261 ms in final position, while the average time spent fixating referents with polysyllabic competitors was 179 ms in medial position and 276 ms in final position. The position of the referent was thus associated with a much larger difference in the time spent fixating the referent when the displayed competitor was polysyllabic (an increase of 97 ms) than when the displayed competitor was monosyllabic (an increase of 24 ms). A two-way repeated measures ANOVA (Position × Type of Competitor)

on the average time spent fixating monosyllabic referents from 200 to 800 ms after the onset of the referent revealed an interaction between Position (medial vs. final) and Type of Competitor (monosyllabic vs. polysyllabic) that was marginally significant by participants ($F_1(1,23) = 3.6$, $p = .07$) and significant by items ($F_2(1,54) = 5.3$, $p < .05$). Planned comparisons revealed a significant difference in the time spent fixating the referent as a function of its position for monosyllabic referents with polysyllabic competitors ($t_1(23) = 3.7$, $p < .001$; $t_2(19) = 4.1$, $p < .001$) but not for monosyllabic referents with monosyllabic competitors. This suggests that the referent was more easily identified in final position than in medial position when the displayed competitor was polysyllabic, but not when the displayed competitor was monosyllabic.

Activation of competitors

Planned comparisons were performed on the average time spent fixating the competitor picture during the entire trial. When the referent was monosyllabic and the competitor monosyllabic, participants spent an approximately equal amount of time fixating the monosyllabic competitor in medial position (142 ms) and final position (145 ms), t_1 and $t_2 < 1$. However, when the referent was monosyllabic and the competitor polysyllabic, participants spent significantly less time fixating the polysyllabic competitor in final position (116 ms) than in medial position (200 ms): $t_1(23) = 3.1$, $p < .005$; $t_2(19) = 3.1$, $p < .005$. The same effect of position, although numerically smaller, was observed when the referent was polysyllabic and the competitor monosyllabic: participants spent significantly less time fixating the monosyllabic competitor in final position (119 ms) than in medial position (150 ms): $t_1(23) = 1.7$, $p = .05$; $t_2(19) = 1.9$, $p < .05$.

In order to examine whether, during the entire trial, the position of monosyllabic referents affected the time spent fixating monosyllabic competitors differently from the time spent fixating polysyllabic competitors, a two-way repeated measures ANOVA with the factors Type of Competitor (monosyllabic vs. polysyllabic) and Position (medial vs. final) was performed on the average time spent fixating the competitor during an entire trial. Type of Competitor was a within-participants variable in the analysis by participants and a between-items variable in the analysis by items. The analysis revealed that the interaction between Position and Type of Competitor was significant by participants ($F_1(1,23) = 4.4$, $p < .05$) but not by items ($F_2(1,54) = 2.9$, $p = .09$). This suggests that during the entire trial, the position of a monosyllabic referent affected the degree to which the displayed competitor was considered for recog-

nition differently depending on whether the competitor was monosyllabic or polysyllabic.

Between 200 and 800 ms after the onset of the target word (which, for trials with monosyllabic referents, corresponds to the time interval during which the proportion of fixations to monosyllabic or polysyllabic competitors, in both medial and final position, exceeded the proportion of fixations to the distractor) participants spent significantly more time fixating monosyllabic competitors of monosyllabic referents in final position (120 ms) than in medial position (89 ms), $t_1(23) = 1.9$, $p < .05$; $t_2(35) = 2.3$, $p < .05$. However, although during the same time interval a numerically reverse effect was observed, participants did not spend significantly less time fixating polysyllabic competitors of monosyllabic referents in final position (103 ms) than in medial position (120 ms). Between 200 and 600 ms after the onset of the target word (which, for trials with polysyllabic referents, corresponds to the time interval during which the proportion of fixations to the monosyllabic competitor exceeded the proportion of fixations to the distractor), participants spent an approximately equal amount of time fixating monosyllabic competitors of polysyllabic referents in medial position (78 ms) as in final position (80 ms). (A non significant effect of identical magnitude was found when the analysis time window was extended to 800 ms and thus identical to the time interval that was used for monosyllabic referents.)

In order to examine whether, during the time interval from 200 to 800 ms after the onset of the referent, the position of a monosyllabic referent affected the time spent fixating a monosyllabic competitor differently from the time spent fixating a polysyllabic competitor, a two-way repeated measures ANOVA was performed on the time spent fixating competitors of monosyllabic referents between 200 and 800 ms after target-word onset. Type of Competitor was a within-participants variable in the analysis by participants and a between-items variable in the analysis by items. The analysis revealed that the interaction between Position (medial vs. final) and Type of Competitor (monosyllabic vs. polysyllabic) was significant ($F_1(1,23) = 4.2$, $p = .05$; $F_2(1,54) = 4.4$, $p < .05$). This suggests that during the time interval when fixations to the competitor indicated that it was considered for recognition, the position of a monosyllabic referent affected the degree to which the displayed competitor was considered for recognition differently depending on whether the competitor was monosyllabic or polysyllabic.

DISCUSSION

The results of Experiment 3.1 demonstrate that the position of a referent in an utterance affects its identification as well as the activation of words that compete with it for recognition.

The analysis of target fixation latencies suggests that the identification of a referent is affected by two factors. First, referents were identified more rapidly in final position than in medial position, across all experimental conditions (i.e., regardless of the syllabic structure of the referent and of its competitor). Second, the identification of monosyllabic referents in final position was facilitated more strongly, compared to the identification of the same referent in medial position, when the competitor was polysyllabic than when the competitor was monosyllabic. The latter finding, though numerically large, was not supported by the statistical analysis, given high variability in the target fixation measure. The pattern of results nevertheless suggests that target fixation latencies for monosyllabic referents reflected the activation of monosyllabic and polysyllabic competitors, and that the activation of these competitors was affected differently by the position of the referent.

This interpretation of the results was supported by analyses that examined the degree to which those competitors were considered for recognition by computing the time spent fixating competitors. The analyses of target fixation latencies suggested that the effect of position on the identification of target words was affected by changes in activation of the displayed competitor as a function of the position of the target word. However, the identification of the target word is influenced by competition from a large number of competitors, not just the displayed competitor. An effect of the position of the referent on the activation of a displayed competitor should therefore be reflected more clearly and directly in fixations to the competitor. Analyses on the time spent fixating the competitor as a function of the position of the referent showed that the position of a monosyllabic referent affected the time spent fixating monosyllabic and polysyllabic competitors differently. When competitor fixations during the entire trial were taken into account, the position of the referent affected the time spent fixating polysyllabic competitors (i.e., they were fixated more in medial than in final position) but not the time spent fixating monosyllabic competitors. When only competitor fixations between 200 and 800 ms after target-word onset were taken into account, the position of the referent affected the time spent fixating monosyllabic competitors (i.e., they were fixated more in final than in medial position) but not the time spent fixating polysyllabic competitors. Crucially, however, in both types of analysis, the position of the monosyllabic referent was shown to affect the degree of activation of monosyllabic and polysyllabic competitors differently.

The effect of the position of the referent on target fixations (see Figures 3-2, 3-3, and 3-4) over time is thus best explained by assuming that target fixations reflect two separate effects. First, accented target words (in final position) tend to be processed more rapidly than unaccented target words (in medial position). Therefore, target fixations tend to increase more rapidly in final than in medial position. However, target fixations are also affected by changes in competitor activation as a function of the position of a monosyllabic target word. When the competitor is monosyllabic, its activation increases in final position, compared to medial position, as a result of which fixations to the target are delayed. Taken together, the effect of accent and the effect of competitor activation would predict a modest effect of position on fixations to a monosyllabic referent when the competitor is monosyllabic. However, when the competitor is polysyllabic, its activation reduces in final position, compared to medial position. As a result of this reduced competition, the target can be identified more quickly in final than in medial position. In combination with the effect of accent that is associated with target position, which also predicts that the target can be identified more quickly in final than in medial position, this would predict that target fixations to monosyllabic referents with polysyllabic competitors would rise much more quickly in final than in medial position.

The analyses on the effect of the position of a monosyllabic referent on the time spent fixating monosyllabic and polysyllabic competitors revealed that the position of the referent did indeed affect differently the degree to which each type of competitor was fixated. This pattern was found regardless of whether the analysis time window extended from 200 ms after target-word onset until the end of the trial or from 200 to 800 ms after the onset of the target word. However, although for both time windows, a numerical *increase* in fixation time for monosyllabic competitors in final position and a numerical *decrease* in fixation time for polysyllabic competitors were observed, the effect of position on fixation time for monosyllabic competitors was only significant over the shorter time window, while the effect of position on fixation time for polysyllabic competitors was only significant over the longer time window. The apparent absence of an effect of the position of a monosyllabic referent on the time spent fixating the monosyllabic competitor during a trial is surprising, given that an inspection of competitor fixation proportions over time (see Figure 3-2) reveals that the position of the monosyllabic referent had a strong effect on fixation proportions to monosyllabic competitors between 200 and 800 ms after the onset of the target word, while competitor and distractor fixations had merged, in both medial and final position, around 800 ms.

A closer inspection of competitor fixation proportions over time, across all three conditions of Experiment 3.1, reveals however that competitor fixation proportions decreased more slowly, after reaching a peak, in medial position than in final position. As a result of this, from around 800 ms after the onset of the target word, fixation proportions to the competitor in medial position tended to be higher than fixation proportions to the competitor in final position. The fact that this pattern was observed after competitor and distractor fixations had clearly merged, and the fact that the same pattern was observed when comparing distractor fixation proportions in medial and final position, indicates that later in the trial, participants had a stronger tendency to fixate pictures other than the target in medial position than in final position. However, such a late effect of position on competitor fixations is unlikely to reflect differences in competitor activation as a function of the position of the target word. Therefore, the time spent fixating a competitor during the entire trial may not adequately reflect the degree to which the competitor was considered for recognition because the magnitude of differences in the time spent fixating competitor pictures as a function of the position of the referent may have been modulated by factors that are not associated with the competitor's lexical activation. However, on the reasonable assumption that such factors would have a similar effect on fixations to monosyllabic and polysyllabic competitors, the two-way analysis which revealed an interaction between the position of the monosyllabic referent and the time spent fixating monosyllabic and polysyllabic competitors clearly demonstrates that the position of the referent affected differently the degree to which monosyllabic and polysyllabic competitors were considered for recognition.

Why were there more fixations to competitor and distractor pictures, especially later in the trial, when the referent occurred in medial position than when it occurred in final position? A possible explanation lies in the relative timing of stimulus events. In Experiment 3.1, each visual display appeared on the computer screen approximately 500 ms before the instruction sentence was presented. However, the referent occurred earlier in the instruction sentence in utterance-medial position than in utterance-final position. The part of the instruction sentence that preceded the referent was therefore of shorter duration when the referent occurred in utterance-medial position ("Put the ...") than when it occurred in utterance-final position ("Now click on the ..."). When participants heard the referent, they had therefore had less time to familiarize themselves with the pictures in the visual display in medial position than in final position. It is possible that the relatively short delay between the appearance of the visual display and the presentation of the referent in medial position encouraged participants to spend more time fixating pictures other than the target picture in me-

dial than in final position, especially later in the trial. Throughout an entire trial, this would act to reduce an (early) increase in the time spent fixating monosyllabic competitors of monosyllabic referents in final position compared to medial position. It would also amplify a decrease in the time spent fixating polysyllabic competitors of monosyllabic referents in final position compared to medial position. The total time spent fixating a competitor during a trial thus may have been influenced by an effect that is associated with the relative timing of stimulus events. This potential confound was addressed by a small change in the design in Experiment 3.2.

Analyses on the time spent fixating competitors over the time interval from 200 to 800 ms after the onset of the target word also found that the position of a monosyllabic referent affected the time spent fixating monosyllabic and polysyllabic competitors differently. This suggests that the effect is not contingent on competitor fixations that occur late in the trial, which are more likely to be influenced by factors other than the lexical activation of the competitor than fixations occurring earlier in the trial. However, although the analyses on the time spent fixating the competitor revealed a predicted and significant *increase* in the time spent fixating monosyllabic competitors of monosyllabic referents in final position compared to medial position, the predicted *decrease* in the time spent fixating polysyllabic competitors of monosyllabic referents, though numerically large, was not significant.

Finally, consider the data when the referent was polysyllabic. Target fixation proportions over time to polysyllabic referents with monosyllabic competitors (see Figure 3-4) appear to show an effect of position on target fixations that is intermediate between the weak effect on targets that was observed for monosyllabic referents with monosyllabic competitors (see Figure 3-2) and the strong effect on targets that was observed for monosyllabic referents with polysyllabic competitors (see Figure 3-3). This suggests that the difference in target fixation proportions (as well as a difference in target fixation latencies) for polysyllabic referents as a function of the position of the referent reflects an effect of the accent associated with the target's position (unaccented in medial position and accented in final position) in combination with a very small or no effect of position on the degree to which the monosyllabic competitor competed for recognition. The analyses of competitor fixations do not provide an unequivocal complimentary picture. The analysis on the total time spent fixating monosyllabic competitors of polysyllabic referents suggests reduced competition of monosyllabic competitors in final position. The analysis on fixations that occur early in the trial, however, shows a small, numerically reversed and non significant effect, suggesting that the position of a polysyllabic referent does not affect the degree to which a monosyllabic competitor is considered for recognition. Taken together, this

suggests that the degree to which a monosyllabic competitor of a polysyllabic referent competes for recognition is not strongly affected by the position of the referent.

EXPERIMENT 3.2

Experiment 3.2 aimed to replicate the finding that prosodically-conditioned variation in the realization of a monosyllabic spoken word has a different impact on the degree of activation of monosyllabic and polysyllabic competitors. Experiment 3.1 demonstrated that prosodic variation in the realization of a monosyllabic referent affects the activation of monosyllabic and polysyllabic competitors differently. A comparison of fixations to competitors when a monosyllabic target word occurred in medial position and when the same word occurred in final position suggested that the processing of a monosyllabic referent in utterance-final position was associated with an *increase* in activation of a monosyllabic competitor and a *decrease* in activation of a polysyllabic competitor. However, these effects were observed on separate trials, since target words were associated either with a monosyllabic competitor or with a polysyllabic competitor. Experiment 3.2 aimed to extend Experiment 3.1, by using a potentially more powerful design in which each monosyllabic referent is associated with a monosyllabic competitor as well as a polysyllabic competitor. This allows one to examine the effect of prosodically-conditioned variation in the realization of a particular monosyllabic referent on the degree to which a monosyllabic and a polysyllabic competitor associated with that particular referent are considered for recognition.

In Experiment 3.1, referents in final position were identified more rapidly than referents in medial position. An additional goal of Experiment 3.2 was to demonstrate that this effect was not contingent on the difference in accent patterns in Experiment 3.1, where referents in medial position were unaccented while referents in final position were associated with a pitch accent. Therefore, in Experiment 3.2, the referent was associated with a pitch accent in medial as well as in final position. Another potential confound in Experiment 3.1 that may have affected the results concerns the fact that, prior to the realization of the referent in the instruction sentence, participants had more time to familiarize themselves with the pictures in the visual display when the referent occurred in medial position (and thus relatively early in the sentence) than when it occurred in final position (and thus relatively late in the sentence). This bias was removed in Experiment 3.2 by using the same relative timing of stimulus events in medial and final condition.

METHOD

Participants

Thirty students of the University of Rochester took part in the experiment. They were all native speakers of American English and were paid a small amount for their participation. None of them had participated in Experiment 3.1.

Materials

Sixteen triplets were constructed using the CELEX lexical database (Baayen et al., 1995). Each triplet consisted of a monosyllabic target word (e.g., *cap*), a monosyllabic competitor (e.g., *cat*) and a polysyllabic competitor (e.g., *captain*). The monosyllabic competitor diverged from the target word at its final segment, which always had the same voicing status as the final consonant of the target word. Vowels preceding voiced consonants tend to be of longer duration than vowels preceding voiceless consonants. Therefore, in order to maximize the degree to which the monosyllabic competitor resembled the target word, it was important that the voicing status of their final consonants was identical. The polysyllabic competitor word had the target word phonemically embedded at its onset. Each triplet was associated with a phonologically unrelated polysyllabic distractor. In addition to the 16 experimental stimulus sets (which are listed in Appendix B), 54 filler sets were constructed. To discourage participants from developing expectations that, in a display comprising pictures with similar names, a monosyllabic word was likely to be the target, 12 of the filler trials comprised three words that started with the same segments (e.g., *bull*, *book*, *bullet*, with one of the monosyllabic words embedded at the onset of the polysyllabic word), of which the polysyllabic word was the target. The remaining 42 filler trials consisted of four phonologically unrelated items. A total of 280 pictures [(16+54 trials) x 4 pictures] representing the four words from each set were selected from the same picture databases that were used for Experiment 3.1.

For each experimental trial, two instruction sentences were constructed using the same sentence frames as in Experiment 3.1. In one instruction sentence, the target word occurred in utterance-medial position (e.g., "Put the cap next to the square") while in the other instruction sentence, the target word occurred in utterance-final position (e.g., "Now click on the cap"). The same sentence frames were used to construct instruction sentences for the 54 filler sets, with the target word occurring in utterance-medial position in half of these sentences and in utterance-final position in the other half of the sentences. All sentences were recorded in a quiet room using a head-mounted microphone. The female speaker (the same speaker as in Experiment 3.1) read the sentences in random order. She was instructed to read the sentences us-

Table 3-4. Mean duration (in ms) of the segments of the monosyllabic target word in the utterance-medial (e.g., "Put the cap next to the square") and utterance-final (e.g., "Now click on the cap") condition of Experiment 3.2.

	Utterance-medial	Utterance-final	Difference	Lengthening
Onset	67	66	-1	-1%
Nucleus	137	165	28	20%
Coda	100	166	66	66%
Total	304	397	93	31%

ing a natural prosodic phrasing of her choice, as long as this phrasing was consistently used for each type of instruction sentence. This resulted in a strong pitch accent on the first syllable of target words in final position and a weaker pitch accent on the first syllable of target words in medial position.

All sentences were digitized and labeled using a speech editor, and durational measurements were made on the target word in utterance-medial and utterance-final position. Table 3-4 presents the average duration of the onset, nucleus and coda of the target words, in utterance-medial and utterance-final position. The target word was markedly longer in utterance-final position (397 ms) than in utterance-medial position (304 ms), an increase of 31%.

Design

Experiment 3.2 consisted of a total of 70 trials (16 experimental trials and 54 filler trials), which consisted of the presentation of the pictures associated with each stimulus set along with an instruction sentence. Two lists were constructed by varying which of the two sentences that had been recorded for every experimental stimulus set was presented along with the visual display. Within each list, the referent occurred in utterance-medial position in half of the experimental trials and in utterance-final position in the other half. Three random orders were created for the two lists, with the constraint that there were never more than two consecutive experimental trials. A set of three filler trials was presented at the beginning of the experiment to familiarize participants with the task and procedure. Fifteen participants were randomly assigned to each list, of which five were assigned to each of the randomizations.

Procedure

The experimental procedure was identical to that of Experiment 3.1, except for a few small changes. The speech files containing the instruction sentences began with a pe-

riod of silence such that there was always a total of 750 ms from the onset of the speech file to the onset of the target word. The experimenter initiated a trial by pressing simultaneously the spacebars of the computer controlling the presentation of the visual display and the computer controlling the presentation of the spoken instruction sentence. The visual display was therefore always presented approximately 750 ms before the onset of the target word. Two further changes to the experimental procedure were aimed to increase participants' familiarity with the visual stimuli. First, prior to the eye-tracking experiment, participants were familiarized with the pictures to ensure that they identified and labeled each picture as intended. (This picture-preview phase was also motivated by the fact that Experiment 3.2 comprised a relatively small number of 8 items per experimental condition.) Each picture appeared on the computer screen along with its printed name and participants pressed the spacebar on the computer keyboard to advance to the next picture. Second, during the experiment, instruction sentences were not preceded by a "Look at the cross" sentence. Instead, participants received instructions prior to the eye-tracking experiment to move the mouse cursor to the central fixation cross at the end of each trial.

Coding Procedure

The coding procedure was identical to that used in Experiment 3.1.

RESULTS AND DISCUSSION

A few trials were discarded because of technical failure or track loss (4 out of 480 trials; 0.8% of the data). On 11 trials, participants erroneously moved or clicked on one of the competitor pictures or moved the target picture without fixating it. These trials (2.3% of the data) were excluded from the analyses.

Figures 3-5 (medial condition) and 3-6 (final condition) present the fixation proportions to the target, the monosyllabic competitor, the polysyllabic competitor, and to the distractor from 0 to 1000 ms after the onset of the target word. Figure 3-7 presents the proportions of fixations to the target in medial and final conditions. In both conditions, fixation proportions to the target started to increase around 300 ms after target-word onset. Target fixations between conditions were comparable until around 700 ms, when fixation proportions to the target picture started to increase more rapidly in final than in medial position. There was a major effect of position on fixations to monosyllabic and polysyllabic competitors. In utterance-medial position, fixations proportions to monosyllabic competitors started to rise around 300 ms after the onset of the target word, and merged with distractor fixations around 700 ms, while fixation

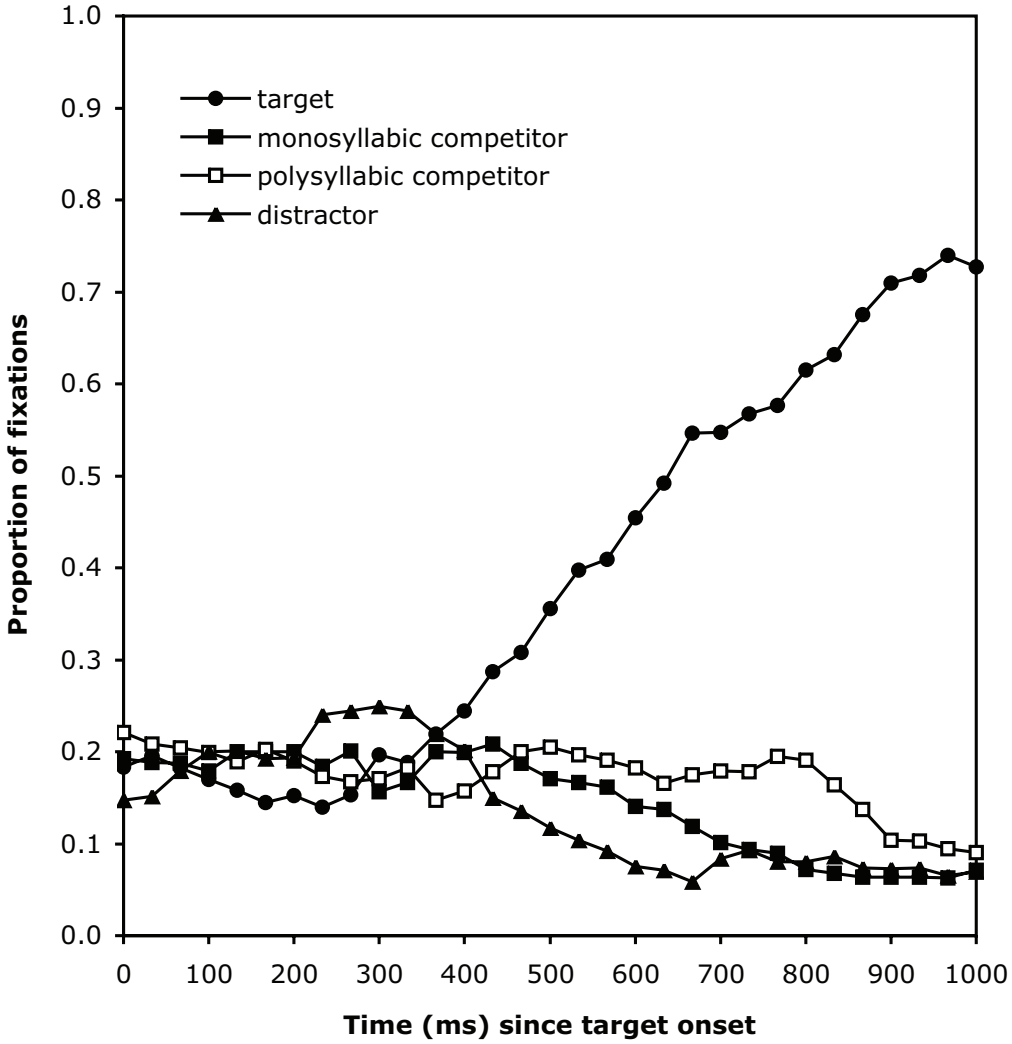

Figure 3-5. Proportion of fixations over time to the target, the monosyllabic competitor, the polysyllabic competitor and the distractor, in the utterance-medial condition of Experiment 3.2.

proportions to polysyllabic competitors started to rise around 400 ms and merged with distractor fixations shortly after 900 ms. Fixation proportions to polysyllabic competitors were higher than fixation proportions to monosyllabic competitors from around 500 to 900 ms after the onset of the target word. In contrast, in utterance-final position, fixation proportions to monosyllabic competitors exceeded fixation proportions to polysyllabic competitors from as early as 200 ms after the onset of the target word until around 800 ms. Fixation proportions to polysyllabic competitors merged with distractor fixation proportions around 600 ms after the onset of the target word, while fixation proportions to monosyllabic competitors merged with distractor fixations shortly before 900 ms.

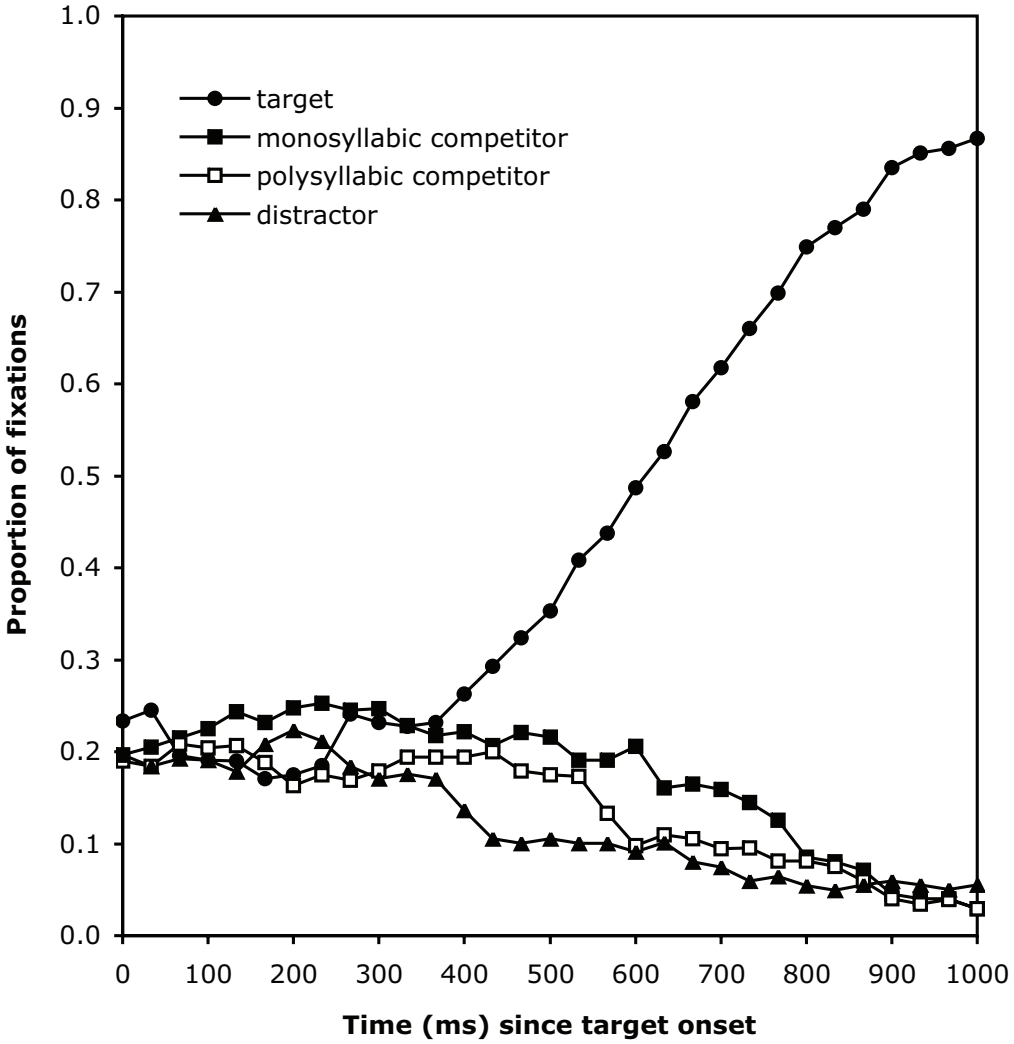

Figure 3-6. Proportion of fixations over time to the target, the monosyllabic competitor, the polysyllabic competitor and the distractor, in the utterance-final condition of Experiment 3.2.

Target identification

To estimate the speed with which the target word was recognized, target fixation latencies were computed across participants and items. The average target fixation latency was shorter when the target word occurred in final position (730 ms) than when it occurred in medial position (979 ms); $t_1(29) = 5.2$, $p < .001$, $t_2(15) = 5.9$, $p < .001$, indicating that the target word was identified more rapidly in final position than in medial position.

Analysis of target fixations over time

In accordance with the criterion that was used to establish the time window for the analysis of the data of Experiment 3.1, an analysis time window was selected that ex-

tended until fixation proportions to each type of competitor, in medial and final position, had merged with distractor fixations. This window extended from 200 to 900 ms after the onset of the target word. The fact that this window was therefore of slightly longer absolute duration than the window that was used in the analyses of Experiment 3.1 (which extended from 200 to 800 ms) may reflect differences between the materials that were used in the two experiments as well as the fact that sentences in Experiment 3.2 were pronounced more slowly than in Experiment 3.1. A planned comparison on the average time spent fixating the referent during this time window revealed that participants spent more time fixating the referent when it occurred in final position (304 ms) than when it occurred in medial position (272 ms; $t_1(29) = 1.7$, $p < .05$; $t_2(15) = 1.9$, $p < .05$). The analyses on target fixation latencies and the time spent fixating the target over time thus converge by showing that referents were identified more rapidly in final than in medial position.

Activation of competitors

To estimate the degree to which monosyllabic and polysyllabic competitors were considered for recognition upon hearing the target word, the total time spent fixating each type of competitor from 200 ms after the onset of the target word up to the end of the trial was computed.

In Experiment 3.2, each experimental display included a picture representing a monosyllabic competitor as well as a picture representing a polysyllabic competitor. The advantage of this design is that it allows to compare fixations to monosyllabic and polysyllabic competitors that were initiated in response to the same acoustic realization of the target word, in medial and final position. However, because eye movements to each type of competitor were recorded at the same time, that is, during the same trial, the degree to which one type of competitor was fixated is likely to affect the degree to which the other type of competitor was fixated. Fixations to each type of competitor were therefore not observed independently, such that paired comparisons between fixations to the two types of competitors within the same display would violate the test's assumption of independence of observations. In order to allow for a proper analysis of the data, two competitor ratios were computed for every participant and every item: one for medial position and one for final position. This ratio expresses the relative degree to which monosyllabic and polysyllabic competitors were fixated, on the basis of the total time spent fixating each competitor during a trial:

$$\frac{\text{t(monosyllabic competitor)}}{\text{t(monosyllabic competitor)} + \text{t(polysyllabic competitor)}}$$

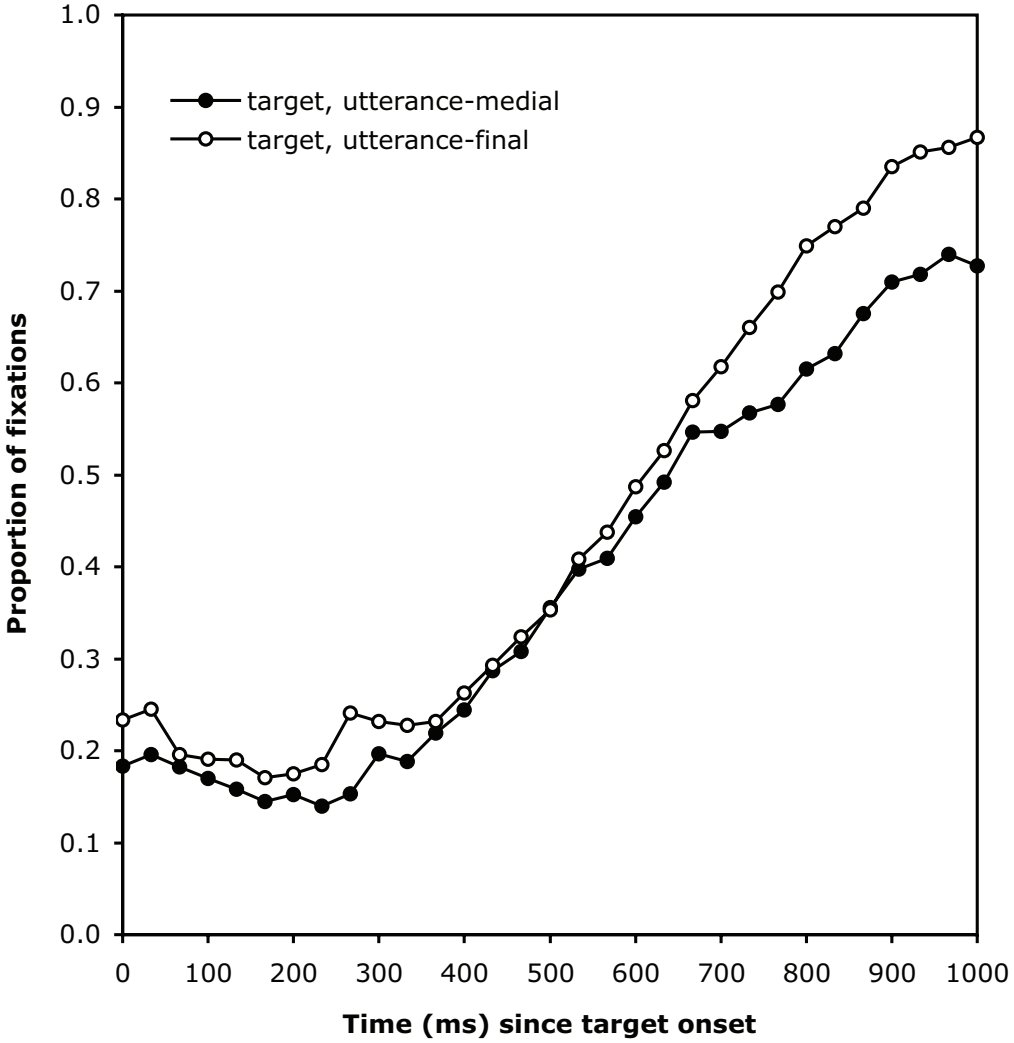

Figure 3-7. Proportion of fixations over time to the target in the utterance-medial and utterance-final condition of Experiment 3.2.

When the target word occurred in medial position, participants spent more time fixating polysyllabic competitors (178 ms) than monosyllabic competitors (147 ms), a competitor ratio of 0.45. When the target word occurred in final position, participants spent more time fixating monosyllabic competitors (163 ms) than polysyllabic competitors (111 ms), a competitor ratio of 0.60. A planned t-test revealed that the competitor ratio was significantly affected by the position of the target word ($t_1(29) = 3.4$, $p < .005$; $t_2(15) = 2.5$, $p < .05$). This demonstrates that the position of the target word affected the relative degree to which monosyllabic and polysyllabic competitors were considered for recognition.

Further planned t-tests examined whether the change in competitor ratio as a function of the position of the target word reflected a significant effect of position on the time spent fixating monosyllabic competitors as well as on the time spent fixating

polysyllabic competitors. Each of these comparisons is only concerned with fixations to one type of competitor, comparing the time spent fixating the competitor in medial and final position, and therefore does not violate the independence of observations assumption. Although participants spent more time fixating a monosyllabic competitor when the target word occurred in final position than when the same target word occurred in medial position (163 ms vs. 147 ms), the effect of position on the time spent fixating monosyllabic competitors was not significant ($t_1(29) = 0.9$, $p = 0.20$; $t_2(15) = 0.6$, $p = 0.26$). However, participants did spend significantly less time fixating a polysyllabic competitor when the target word occurred in final position than when the same target word occurred in medial position (111 ms vs. 178 ms; $t_1(29) = 3.3$, $p < .005$, $t_2(15) = 3.4$, $p < .005$).

Planned t-tests were also performed on the average time spent fixating each type of competitor over the time interval from 200 to 900 ms after target-word onset. During this time interval, participants spent more time fixating a monosyllabic competitor when the target word occurred in final position (143 ms) than when the same target word occurred in medial position (103 ms). A planned comparison revealed that the effect of the position of the target word on the time spent fixating monosyllabic competitors was significant: $t_1(29) = 2.9$, $p < .005$, $t_2(15) = 2.5$, $p < .05$. During the same time interval, participants spent less time fixating a polysyllabic competitor when the target word occurred in final position (98 ms) than when the same target word occurred in medial position (124 ms). A planned comparison again revealed that the effect of the position of the target word on the time spent fixating polysyllabic competitors was significant: $t_1(29) = 1.9$, $p < .05$, $t_2(15) = 2.2$, $p < .05$. This effect was also supported by a difference in competitor ratio (i.e., the relative degree to which monosyllabic and polysyllabic competitors were fixated) as a function of the position of the target word. The competitor ratio was 0.47 in medial position and 0.61 in final position. A planned comparison revealed that this difference was significant: $t_1(29) = 3.4$, $p < .005$; $t_2(15) = 3.2$, $p < .005$. These analyses clearly demonstrate that monosyllabic competitors competed for recognition more strongly in final position than in medial position, while polysyllabic competitors competed for recognition more strongly in medial than in final position.

GENERAL DISCUSSION

This study explored whether variation in the acoustic realization of the speech signal associated with the position of a word in a spoken utterance affects the identification of that word. In particular, it compared the processing of referents in utterance-medial

and utterance-final position. In order to assess the identification of the referent and the transient activation of competitors associated with the referent, we presented participants with a visual display consisting of four pictures and concurrent spoken instructions referring to one of the pictures. Each visual display included a picture representing the referent as well as one or two pictures representing competitor words whose name(s) overlapped with the initial sounds of the target word. The lexical activation of the referent and of its competitor(s) was estimated from participants' eye movements to the pictures in the visual display, as the name of the referent was heard and processed. The referent occurred in utterance-medial position, as in "Put the cap next to the square", and in utterance-final position, as in "Now click on the cap". In accordance with our expectations based on the phonetics literature, the acoustic realization of the referent was strongly affected by this manipulation. The referent tended to be of longer duration in utterance-final position than in utterance-medial position. Furthermore, and consistent with the literature, we observed that final lengthening most strongly affected the realization of segments immediately preceding the utterance boundary, thus affecting the initial segments of monosyllabic words much stronger than the initial segments of polysyllabic words. Of interest was whether fixations to the referent, and transient fixations to the competitor, would be affected by these durational differences or associated acoustic differences that reflect the referent's position in the utterance.

This study revealed two main findings, which taken together make an important contribution to our understanding of the processes involved in the recognition of spoken words. First, the results of Experiment 3.1 and 3.2 converge by demonstrating that the position of a spoken word in an utterance has a systematic impact on its identification. Across all conditions of the experiments, the latency to fixate the picture of the referent was shorter when the referent occurred in final position than when it occurred in medial position. The fact that this effect was observed when referents were deaccented in medial position and associated with a pitch accent in final position (in Experiment 3.1), as well as when referents in medial as well as final position were associated with a pitch accent (in Experiment 3.2) suggests that the effect is not contingent on a comparison involving accented versus deaccented referents.

The results of this study do not speak further to the nature of the acoustic information that facilitated the identification of referents in utterance-final position, however. One possibility is that such information is associated with the realization of the referent itself. For instance, a relatively strong pitch accent and final lengthening of the referent in utterance-final position may have enhanced the salience of the phonetic features of the initial sounds of the referent (cf. Bard, 1990; Cho, 2002; Fougeron &

Keating, 1997), thus facilitating the mapping of the speech signal onto lexical representations. Another possibility is that the presence of a relatively strong pitch accent on the initial syllable of the referent in utterance-final position may facilitate the mapping between the speech signal and candidate words by leading listeners to attend to specific phonetic properties of the speech signal (Terken & Nooteboom, 1987). Alternatively, the identification of the referent in utterance-final position may be facilitated by the processing of information preceding the referent. This is because the utterance boundary is likely to affect not only the realization of the referent, but also, at least to some degree, the realization of segments preceding the referent. Listeners may have used this information (e.g. pitch declination towards the end of the utterance, or lengthening of the segments immediately preceding the referent) to focus their attention in anticipation of the last word of the utterance (cf. Cutler, 1976; Pitt & Samuel, 1990). It is likely that both of these factors facilitate the identification of referents in final position. The crucial point, however, is that each of these factors, to the extent that they had an impact on the recognition process, concern speech variation which is conditioned by prosodic structure.

The present study did not manipulate independently the degree of pitch accent associated with the referent and its position in the utterance. Since both of these aspects of prosodic structure are manifested in the speech signal by, amongst other things, the lengthening of speech sounds, it is likely that the lengthening observed in utterance-final position compared to utterance-medial position reflected the acoustic manifestation of the utterance boundary (i.e., constituent-level prosodic structure) as well as the acoustic manifestation of a relatively strong pitch accent associated with the referent in utterance-final position (i.e., prominence-level prosodic structure). It is therefore reasonable to assume that lengthening associated with the pitch accent on the referent and final lengthening associated with the referent's position in the utterance's prosodic structure did not have independent effects on lexical processing. Further research would be needed to establish the degree to which pitch accent and final lengthening each contribute to the effects observed in this study, for example by manipulating the degree of pitch accent associated with a spoken word in utterance-final position.

The second main finding, and the focus of these experiments, is that the position of a spoken word in an utterance can have an impact on the degree of activation of words that compete with it for recognition. In particular, the position of a monosyllabic referent had a different impact on the degree to which words of different syllabic length (i.e., monosyllabic and polysyllabic competitors) were considered for recognition. Upon hearing a monosyllabic referent, participants spent *more* time fixating a

monosyllabic competitor when the referent occurred in final position than when it occurred in medial position. Conversely, they spent *less* time fixating a polysyllabic competitor when the monosyllabic referent occurred in final position than when it occurred in medial position. In Experiment 3.1, where each monosyllabic referent was associated with either a monosyllabic or a polysyllabic competitor, the position of the referent affected the activation of monosyllabic and polysyllabic competitors differently, but separate effects of the position of the referent on the activation of monosyllabic as well as polysyllabic competitors were not always statistically robust. In Experiment 3.2, where each monosyllabic referent was associated with both a monosyllabic and a polysyllabic competitor, a comparison between the degree of activation of those competitors in medial and in final position demonstrated that the processing of the referent in final position was associated with both a significant *increase* in activation of monosyllabic competitors, as well as a significant *decrease* in activation of polysyllabic competitors.

The finding that the pattern of competitor activation that is associated with the processing of a monosyllabic referent varies as a function of the referent's position in an utterance presents a fundamental challenge for many theories and models of spoken-word recognition. Current theories and models generally agree that as a spoken word unfolds, its identification is constrained by competition from similar-sounding candidate words. The models differ, however, in their assumptions about which acoustic information in the speech signal is relevant for the evaluation of candidate words. The predominant view, as propagated by most computational models, is that the architecture and representations of a model should only be sensitive to lexically contrastive information in the speech signal. Several important computational models have made the assumption that such information can be captured and processed effectively by phonemic representations. In the TRACE model of speech perception (McClelland & Elman, 1986), lexical representations consist of a sequence of phonemes, and the degree of support for lexical candidates is computed exclusively on the basis of their phonemic overlap with a phonemic representation of the speech signal. It is clear that such an evaluation mechanism cannot account for the findings of the present study.

Phonemic representations are also central to the original version of the Shortlist model (Norris, 1994). In this model, the support for a candidate word is a function of the degree to which it phonemically matches and mismatches the speech signal. More recent versions of the model (Norris, McQueen, Cutler, & Butterfield, 1997; Norris, McQueen, & Cutler, 2000) have extended its original architecture by adding several components that allow the model to posit word boundaries in a strictly phonemic rep-

resentation of the input. Importantly, these components act on the basis of information that cannot be captured by the model's phonemic representations, such as stored knowledge about phonotactics, or suprasegmental information, which is used to locate the onset of stressed syllables. Candidate words that are misaligned with likely word boundaries are penalized (their activation is halved; see Norris et al., 1997, for details). The model thus indirectly uses information in the speech signal that cannot be captured by its phonemic representations to assist the evaluation of lexical candidates. The current version of the Shortlist model does not, however, include a component that is sensitive to subphonemic variation associated with constituent-level prosodic structure. In order to account for the systematic effect of prosodically-conditioned speech variation on the evaluation of candidate words observed in this study, Shortlist would need to be modified.

One solution would be to extend the current version of the Shortlist model in accordance with a proposal that was made by Salverda et al. (2003). They suggested that listeners compute a prosodic representation of the speech signal. On the basis of aspects of this structure, the word-recognition system develops expectations about the location and strength of prosodic boundaries in the input. Within the framework of the current version of the Shortlist model, aspects of this prosodic structure, such as the edges of prosodic constituents larger than the word, would correspond to the location of likely word boundaries. Such information could be used to assist the evaluation of lexical hypotheses based on phonemic representations by providing a boost in activation to candidate words that were aligned with these likely word boundaries. Candidates that were aligned with these boundaries would therefore be favored in the recognition process.

The findings of this study can also be accommodated within the existing framework of models of spoken-word recognition that incorporate highly detailed lexical representations. For instance, in exemplar-based models (e.g., Goldinger, 1998; Johnson, 1997), lexical representations consist of multiple exemplars of a spoken word and each of these exemplars contains all the acoustic properties of a word. Listeners have acquired such representations on the basis of their experience with the speech input. The lexical representation of each spoken word therefore includes different realizations of that word across different prosodic contexts. This allows exemplar-based models to account for the impact of naturally-occurring prosodically-conditioned variation in the realization of a spoken word on the activation of competitors in a natural way. The initial sounds of a monosyllabic word in utterance-final position will tend to match existing monosyllabic exemplars (in particular the exemplars that occurred in utterance-final position and thus tend to be affected by this prosodic context

in a similar way, e.g. strongly affected by final lengthening) better than existing polysyllabic exemplars (some of which occurred in utterance-final position, but whose initial sounds were affected by final lengthening less strongly).

As mentioned in the introduction, in phoneme-based models of spoken-word recognition (e.g., TRACE and Shortlist), lexical activation is strongly affected by phonemically contrastive information in the speech signal. According to this view, the degree to which a candidate competes for recognition upon hearing a spoken word is strongly determined by its degree of segmental overlap with the speech input (e.g., in Shortlist, computed in terms of the number of matching and mismatching phonemes). This predicts that the candidate *captain* will *always* be a stronger competitor of the spoken word *cap* than the candidate *cat*, because the word *captain* overlaps with the word *cap* by three phonemes and with the word *cat* by only two phonemes. The results of Experiment 3.2 are, however, clearly inconsistent with this notion. When the word *cap* occurred in utterance-medial position, the candidate *captain* competed for recognition more strongly than the candidate *cat*. This finding is in accordance with the predictions of phoneme-based models. When the word *cap* occurred in utterance-final position, however, the candidate *cat* competed for recognition more strongly than the candidate *captain*, even though the speech signal phonemically matched the candidate *captain* to a greater extent than the candidate *cat* did. This finding suggests that the activation of candidate words can be strongly affected by subphonemic information in the speech signal. Interestingly, it further suggests that the informational value of particular phonetic information (e.g., phonemically contrastive information) is context dependent. In utterance-medial position, the identification of a word may strongly rely on the processing of phonemic information because in this context, phonemic differences between words may tend to be more salient than subphonemic differences. In utterance-final position, however, the realization of words tends to be more strongly marked by subphonemic information conditioned by prosodic structure, increasing the value of such information for the evaluation of lexical candidates. Therefore, final lengthening of the initial sounds of the word *cap* in utterance-final position may render the monosyllabic candidate *cat* (whose initial sounds would be affected by final lengthening in a similar way) a stronger competitor than the polysyllabic word *captain*, even though the latter overlaps with the speech signal for a greater number of segments. In other words, when the spoken sequence /kæp/ is heard, information in the speech signal that is phonemically inconsistent with the competitor *cat* appears to have a bigger impact on the activation of this competitor in medial position than in final position.

There is now a substantial body of research showing that acoustic correlates of prosodic structure can systematically affect lexical activation (Cho, McQueen, & Cox, submitted; Christophe, Peperkamp, Pallier, Block, & Mehler, 2004; Crosswhite, McDonough, Masharov, & Tanenhaus, submitted; Davis et al., 2003; Gow & Gordon, 1995; Salverda et al., 2003; Shatzman & McQueen, in press). Although some prosodically-conditioned phonetic variation may affect the recognition process straightforwardly, for instance by facilitating the processing of spoken words in particular prosodic domains (cf. the finding in both experiments of the present study that words in utterance-final position tend to be identified more rapidly than words in utterance-medial position), the primary contribution of the current study is that it demonstrates that speech variation associated with prosodic structure can act to affect dynamically the degree to which different types of words compete for recognition. This was shown contrasting the lexical activation of competitors when a referent was followed by a prosodic-word boundary (in utterance-medial position) versus when the same referent was followed by an utterance boundary (in utterance-final position). Because the difference in the size of the prosodic boundary following the referent between these two conditions was thus relatively large, we expected that this would result in large durational (and other, prosodically-conditioned) differences between the acoustic realization of referents in utterance-medial and utterance-final position, which would maximize the likelihood of finding a systematic effect of the referent's position on the degree of activation of competitors. These differences in competitor activation could reflect a *gradient* effect of subphonemic variation associated with constituent-level prosodic structure on lexical activation. That is, competitor activation patterns could vary continuously as a function of the nature and size of the prosodic constituents associated with a spoken word.

An alternative, more conservative, interpretation is that the processing of a monosyllabic referent in utterance-final position is associated with a different pattern of competitor activation than the processing of the same referent in any other position in an utterance. That is, the utterance-final position may be special. This may be the case if the realization of the referent in utterance-final position is characterized by specific acoustic cues and if the evaluation of candidate words is sensitive to these cues, for instance the fact that the final segments of a referent in utterance-final position are not coarticulated by following context. Nevertheless, such cues are conditioned by the prosodic structure of the utterance, and even if it could be shown that they had an impact on the evaluation of candidate words in the present study, this would not demonstrate that these cues are necessary to observe such an effect.

Further research is needed to establish whether there are gradient effects of prosodic structure on the relative ranking of competitor words, beyond any effects that may be specific to utterance-final position. It seems reasonable to suppose that this is the case, however, since previous research has provided some evidence for a gradient effect of the strength of a prosodic-word boundary on lexical activation for different realizations of a word in the *same* utterance context. Salverda et al. (2003) showed that the degree to which the initial sounds of a lexically ambiguous fragment (e.g., /hɑm/, which may correspond to the Dutch word *ham*, id., or the onset of the Dutch word *hamster*, id.) were marked by a prosodic-word boundary, as estimated by the duration of the fragment, systematically affected the lexical interpretation of such a fragment. The longer the duration of the fragment, the more its interpretation was biased towards a monosyllabic word (*ham*) as opposed to a longer word (e.g., *hamster*), thus suggesting that the effect of the prosodic boundary on lexical activation was proportional to the degree to which the boundary was realized.

To conclude, the research reported in this study makes a primary contribution to our understanding of the recognition of words in continuous speech by showing that the position of a word in an utterance can affect the pattern of lexical activation associated with the evaluation of candidate words. Systematic variation in the realization of a spoken word that is associated with its position in an utterance had a systematic effect on the speed with which the word was identified as well as on the degree to which different types of competitors were involved in the competition process. When the referent was monosyllabic, competition from monosyllabic competitors *increased* in utterance-final position (compared to processing of the same referent in utterance-medial position), while competition from polysyllabic competitors *decreased*. This demonstrates that naturally occurring phonetic variation conditioned by constituent-level prosodic structure can play a central role in the evaluation of lexical candidates. This study therefore converges with a growing body of research (Cho et al., submitted; Christophe et al., 2004; Crosswhite et al., submitted; Davis et al., 2002; Gow & Gordon, 1995; Gow, 2002; Kemps, 2004; Salverda et al., 2003; Shatzman & McQueen, in press; Spinelli et al., 2003; Tabossi, Collina, Mazzetti, & Zoppello, 2000) in showing that fine-grained phonetic detail in the realization of words in continuous speech is preserved in the representations that mediate the recognition of spoken words, rather than being discarded as irrelevant information. The primary contribution of the present research is that it extends these studies by demonstrating that prosodically-conditioned variation in the realization of words in continuous speech can act to modulate the lexical competition process dynamically, by having a different impact on the evaluation of different types of candidate words. A word that may be a

strong competitor of a spoken word in one utterance position may be a weaker competitor of the same word in another utterance position.

REFERENCES

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

Alphen, P. M. van, & McQueen, J. M. (in press). The effect of Voice Onset Time differences on lexical access in Dutch. *Journal of Experimental Psychology: Human Perception and Performance.*

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. (1990). Competition, lateral inhibition, and frequency: Comments on the chapters of Frauenfelder and Peeters, Marslen-Wilson, and others. In G. E. Altmann (Ed.), *Cognitive models of speech processing* (pp. 183—210). Cambridge, MA: MIT Press.

Beckman, M.E. (1996) The parsing of prosody. *Language and Cognitive Processes, 11*, 17—67.

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.

Campbell, W. N., & Isard, S. D. (1991). Segment durations in a syllable frame. *Journal of Phonetics, 19*, 37—47.

Cho, T. (2002). The Effects of Prosody on Articulation in English. New York: Routledge.

Cho, T., McQueen, J. M., & Cox, E. A. (submitted). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English.

Christophe, A., Peperkamp, S., Pallier, C., Block, E., & Mehler, J. (2004). Phonological phrase boundaries constrain lexical access: I. Adult data. *Journal of Memory and Language, 51*, 523—547.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 130—163). Hillsdale, NJ: Erlbaum.

Crosswhite, K., McDonough, J., Masharov, M., & Tanenhaus, M. K. (submitted). Phonetic cues to word length in the online processing of onset-embedded word pairs. *Journal of Phonetics*.

Cutler, A. (1976). Phoneme-monitoring reaction times as a function of preceding intonation contour. *Perception & Psychophysics, 20*, 55—60.

Cutler, A., & Foss, D. J. (1977). On the role of sentence stress in sentence processing. *Language and Speech, 20*, 1—10.

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

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.

Edwards, J., Beckman, M. E., & Fletcher, J. (1991). The articulatory kinematics of final lengthening. *Journal of the Acoustical Society of America, 89*, 369—382.

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

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

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

Gow, D. W., 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.

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.

Johnson, K. (1997). 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.

Kemps, R. J. J. K. (2004). *Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction*. Doctoral Dissertation, University of Nijmegen (MPI Series in Psycholinguistics, 28).

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.

Kuhl, P. K., Andruski, J. E., Chistovich, I. A., Chistovich, L. A., Kozhevnikova, E. V., Ryskina, V. L., Stolyarova, E. I., Sundberg, U., & Lacerda, F. (1997). Cross-language analysis of phonetic units in language addressed to infants. *Science, 277*, 684—686.

Ladd, D. R., & Campbell, W. N. (1991). Theories of prosodic structure: Evidence from syllable duration. *Proceedings of the XII th International Congress of Phonetic Sciences* (pp. 290—293). Aix-en-Provence, France.

Luce, P. A. (1986). *Neighborhoods of words in the mental lexicon*. Ph.D. dissertation, Indiana University. Technical Report No. 6, Speech Research Laboratory, Department of Psychology, Indiana University.

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The Neighborhood Activation Model. *Ear and Hearing, 19*, 1—36.

Luce, P. A., Pisoni, D. B., & Goldinger, S.D. (1990). Similarity neighborhoods of spoken words. In G. T. M. Altmann (Ed.), *Cognitive models of speech processing: Psycholinguistic and computational perspectives* (pp. 122—147). Cambridge, MA: MIT Press.

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

Marslen-Wilson, W. D. (1993). Issues of process and representation in lexical access. In G.T.M. Altmann & R. Shillcock (Eds), *Cognitive models of speech processing: The second Sperlonga meeting* (pp. 187—210). Hillsdale, NJ: Lawrence Erlbaum Associates Inc.

Marslen-Wilson, W. D., & Welsh, A. (1978). Processing interactions and lexical access during word recognition in continuous speech. *Cognitive Psychology, 10*, 29—63.

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

McMurray, B., Tanenhaus, M. K., & Aslin, R. N. (2002). Gradient effects of within-category phonetic variation on lexical access. *Cognition, 86*, B33—B42.

McQueen, J. M., Dahan, D. & Cutler, A. (2003). Continuity and gradedness in speech processing. In A.S. Meyer & N.O. Schiller (Eds.), *Phonetics and phonology in language comprehension and production: Differences and similarities* (pp. 39—78). Berlin: Mouton de Gruyter.

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

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

Norris, D., McQueen, J. M., Cutler, A., & Butterfield, S. (1997). The possible-word constraint in the segmentation of continuous speech. *Cognitive Psychology, 34*, 191—243.

Norris, D., McQueen, J. M., & Cutler, A. (2000). Merging information in speech recognition: Feedback is never necessary. *Behavioral and Brain Sciences, 23*, 299—370.

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.

Pitt, M. A., & Samuel, A. G. (1990). The use of rhythm in attending to speech. *Journal of Experimental Psychology: Human Perception and Performance, 16*, 564—573.

Salverda, A. P., Dahan, D., & McQueen, J. M. (2003). The role of prosodic boundaries in the resolution of lexical embedding in speech comprehension. *Cognition, 90*, 51—89.

Selkirk, E. (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 investors of auditory sentence processing. *Journal of Psycholinguistic Research, 25*, 193—247.

Shatzman, K. B., & McQueen, J. M. (in press). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*.

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.

Tanenhaus, M. K., Magnuson, J. S., Dahan, D., & Chambers, C. (2000). Eye movements and lexical access in spoken-language comprehension: Evaluating a linking hypothesis between fixations and linguistic processing. *Journal of Psycholinguistic Research, 29*, 557—5580.

Terken, J., & Nooteboom, S. G. (1987). Opposite effects of accentuation and deaccentuation on verification latencies for given and new information. *Language and Cognitive Processes, 2*, 145—163.

Utman, J. A., Blumstein, S. E., & Burton, M. W. (2000). Effects of subphonetic and syllable structure variation on word recognition. *Perception & Psychophysics, 62*, 1297—1311.

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.

APPENDIX A: EXPERIMENT 3.1 STIMULUS SETS

Each item from a target-competitor pair was the target for half of the participants and the competitor for the other half of the participants (and vice versa for the other item from a target-competitor pair). Items with an asterisk were discarded from the analyses.

Monosyllabic referents with monosyllabic competitors

Target-Competitor pairs		Distractors	
back	bat	cocktail	flashlight
beak	beet	teapot	doughnut
bud	bug	cannon	funnel
cap	cat	hurdle	racket
coat	coke	squirrel	trumpet
comb	cone	pepper	statue
foam	phone	bison	champagne
graph	grass	matches	chisel
gum	gun	turtle	beaker
harp	heart	turkey	curtains
leaf	leash	table	blender
map	mat	giraffe	anchor
mouse*	mouth*	bracelet	trophy
neck	net	cigar	mountain
road	robe	canoe	circus
sheep*	sheet*	kiwi	outlet
suit	soup	arrow	lantern
tack	tap	compass	eagle
track	trap	sandwich	button
tub	tug	ferret	mailbox

Monosyllabic referents with polysyllabic competitors (and polysyllabic referents with monosyllabic competitors)

Target-Competitor pairs		Distractors	
ant	antlers	chain	lemon
bee	beetle	chair	guitar
bull	bullet	house	piano
cab	cabin	tree	mixer
can	candy	saw	helmet
car	carpet	pear	lobster
cart	carton	glove	shovel
check	checkers	broom	garlic
doll	dolphin	plane	lighthouse
ham	hamster	door	carrot
knee	needle	swan	zipper
pick	pickle	bulb	towel
pill	pilgrim	kite	tiger
pie	pirate	swing	zebra
pump	pumpkin	fan	medal
rat	rattle	desk	orange
rock	rocket	hook	walnut
sole	soldier	dog	bandage
tie	timer	drill	sausage
toe	toaster	owl	grenade

APPENDIX B: EXPERIMENT 3.2 STIMULUS SETS

The first member of a pair of distractors marked with an asterisk was replaced with the second member of that pair after 6 participants had been tested.

Target	Monosyllabic competitor	Polysyllabic competitor	Distractor
beak	beet	beaker	whistle
bell	bed	bellows	scissors
bowl	bone	boulder	fountain
bug	bud	buggy	shovel
cap	cat	captain	guitar*/beaker
carp	cart	carpet	ladder
doll	dog	dolphin	magnet
leaf	leash	leaflet	cigar
neck	net	nectarine	letter
pad	pan	paddle	bucket
pick	pit	pickle	ribbon
rack	rat	racket	garlic
robe	road	robot	table
tack	tap	taxi	dagger*/lemon
track	trap	tractor	lighter
well	web	welder	feather

GENERAL DISCUSSION

CHAPTER 4

The vast majority of listeners are blissfully unaware of the complex cognitive processes that enable them to comprehend spoken language. One of the challenges that the listener has to face in order to understand the message of the speaker is to recognize the words in the speech stream. This requires the mapping of information extracted from a transient, highly variable and rapidly unfolding speech signal onto stored representations of lexical form. The speech signal contains a plethora of types of acoustic information that could potentially help the listener recognize spoken words. This raises the question of which types of acoustic information are relevant for the recognition of spoken words. The research reported in this dissertation is concerned with the influence on lexical processing of speech variation that is conditioned by constituent-level prosodic structure.

SUMMARY OF RESULTS

The experiments in Chapter 2 examined the processing of spoken sequences that either corresponded to a monosyllabic word (e.g., ham) or to the initial syllable of a polysyllabic word (e.g., hamster) that has the monosyllabic word embedded at its onset. These sequences were thus phonemically identical. However, the sequence's position in the prosodic structure of the utterance was different depending on whether it corresponded to a monosyllabic or a polysyllabic word. This is because the monosyllabic word was followed by a prosodic-word boundary, but the initial syllable of a polysyllabic word was not (and indeed could never be). On the basis of the phonetics literature, it was expected that the acoustic realization of the sequence would be affected differently depending on whether it was aligned at offset with a prosodic-word boundary. In line with these expectations, it was found that sequences that corresponded to a monosyllabic word (e.g., *ham*) tended to be of longer duration than sequences that were phonemically identical but that corresponded to the initial syllable of a polysyllabic word (e.g., *ham* in *hamster*). Of interest was whether such prosodically-conditioned differences between the acoustic realizations of the initial sounds of monosyllabic and polysyllabic words would affect listeners' lexical interpretation of the sequences.

In the experiments presented in Chapter 2, Dutch listeners heard spoken sentences including a polysyllabic word (e.g., the word *hamster* (id.) in *Ze dacht dat die hamster verdwenen was*, she thought that that hamster had disappeared). The first syllable of the polysyllabic word had been replaced by a recording of a monosyllabic word (*ham*, id.) or by the initial syllable of another recording of the polysyllabic word. The syllable was of longer duration when it corresponded to a monosyllabic word than when it corresponded to the initial syllable of a polysyllabic word. Listeners were presented with a visual display on a computer screen including a picture representing the monosyllabic word and a picture representing the polysyllabic word. They had been instructed to move, with the computer's mouse, the picture corresponding to the word that was mentioned in the spoken sentences (i.e., the picture representing the polysyllabic word, e.g. hamster in *Ze dacht dat die hamster verdwenen was*). Throughout the experiment, participants' eye movements were recorded. Listener's interpretation of the initial syllable of the cross-spliced target word (e.g., *hamster*), as the speech signal unfolded, was assumed to be reflected by the degree to which listeners fixated the pictures representing the monosyllabic and the polysyllabic word.

Experiment 2.1 provided strong evidence that the lexical interpretation of the first syllable of the cross-spliced carrier word was affected by subphonemic information in the speech signal. There were more looks to the picture representing the monosyllabic word when the first syllable of the polysyllabic target word originated from a recording of a monosyllabic word than when it originated from a different recording of the polysyllabic word. This effect was large and statistically significant in Experiment 2.1A, in which the monosyllabic word had been recorded in a sentence where it was followed by a stressed syllable, while the effect was smaller and not statistically significant in Experiment 2.1B, in which the monosyllabic word had been recorded in a sentence where it was followed by an unstressed syllable. This suggests that the acoustic cues that assisted the interpretation of the initial syllable of the target word as a monosyllabic word or as the first syllable of a polysyllabic word were subject to variability. Indeed, the average durational difference between the monosyllabic word and the initial syllable of the polysyllabic word was larger in Experiment 2.1A (20 ms) than it was in Experiment 2.1B (15 ms).

On the basis of the results of Experiment 2.1, it was hypothesized that listeners favored an interpretation of the initial syllable of the cross-spliced polysyllabic target word as corresponding to a monosyllabic word to the degree that the acoustic realization of the sequence was associated with lengthening, and therefore characterized by speech variation that is associated with a following prosodic boundary. This interpretation of the data received additional support from the finding that across all the items

that were used in the experiment, there was a significant correlation between the difference in duration of the initial syllable of the cross-spliced target word when it originated from a monosyllabic word versus when it originated from a polysyllabic word, and the difference in the degree to which listeners fixated the picture representing the monosyllabic word upon hearing the different versions of the cross-spliced target word associated with these sequences.

Experiments 2.2 and 2.3 extended the findings of Experiment 2.1 and tested predictions generated by the prosodic-boundary hypothesis. Cross-spliced target words were created in which the duration of the target word's first syllable was manipulated in a systematic way. This was done by selecting, from the original set of sentences that had been recorded for Experiment 2.1, monosyllabic words and initial syllables of polysyllabic words on the basis of their duration.

In Experiment 2.2, cross-spliced target words were created such that the target word's initial syllable was of approximately equal duration when it originated from a monosyllabic word as when it originated from a polysyllabic word. It was predicted that, if the duration of the ambiguous syllable reflects the degree to which it is associated with a following prosodic boundary, and if this information is used by listeners to guide their lexical interpretation of the sequence, the origin of the initial syllable of the cross-spliced target word should not affect listeners' lexical interpretation of the sequence. This prediction was confirmed by the finding that the degree to which listeners fixated the picture representing the monosyllabic word did not vary as a function of whether the initial syllable of the polysyllabic target word originated from a monosyllabic word or from the initial syllable of a polysyllabic word.

Experiment 2.3 used cross-spliced target words that had been created such that the initial syllable of the target word was of longer duration when it originated from a polysyllabic word than when it originated from a monosyllabic word. In the sentences that were recorded for the experiments reported in Chapter 2 as well as in the Dutch language in general, monosyllabic words tend to be of longer duration than the initial syllables of polysyllabic words that have these monosyllabic words embedded at their onset. In Experiment 2.3, the duration of the initial syllable of the cross-spliced target word was therefore more characteristic of a monosyllabic word when the syllable originated from a polysyllabic word (and was of relatively long duration) than when it originated from a monosyllabic word (and was of relatively short duration). This was reflected by listeners' lexical interpretation of the first syllable of the target word, as estimated from their eye movements. There were more fixations to the picture representing the monosyllabic word when the first syllable of the target word was of relatively long duration and originated from a polysyllabic word than when the first sylla-

ble of the target word was of relatively short duration and originated from a monosyllabic word. Taken together, the results of the experiments presented in Chapter 2 demonstrate that listeners' interpretation of a spoken sequence that is phonemically fully ambiguous between a monosyllabic word and the initial syllable of a polysyllabic word is affected by the duration of the sequence more than by the word that this sequence originates from. This suggests that the acoustic cues that affected listeners' interpretation of the first syllable of the polysyllabic target word are not invariantly associated with the production of monosyllabic and polysyllabic words but rather associated with and dependent on the acoustic manifestation of prosodic-word boundaries.

The experiments reported in Chapter 2 demonstrate that listeners can distinguish a sequence that corresponds to a monosyllabic word from a sequence that is phonemically identical and that corresponds to the initial syllable of a polysyllabic word. The difference between these two sequences is that the sequence is followed by a prosodic boundary when it corresponds to a monosyllabic word but not when it corresponds to a polysyllabic word. When this prosodic boundary is phonetically realized, that is, when a monosyllabic word is of relatively long duration, such information can act to assist the listener's lexical interpretation of the unfolding speech signal.

The goal of Chapter 3 was to extend the findings of Chapter 2 by contrasting the processing of a word across different, prosodically-defined positions in an utterance. The acoustic realization of a word is affected by its position in the prosodic structure of an utterance. Variation in the realization of a particular word that is associated with its position in an utterance may therefore affect lexical processing. Chapter 3 examined whether the ease with which a word can be identified and the degree to which processing of the word is associated with the consideration for recognition of different types of similar-sounding words varies as a function of the word's position in an utterance.

The experiments reported in Chapter 3 contrasted the recognition of words in utterance-medial (e.g., the word *cap* in *Put the cap next to the square*) and utterance-final position (e.g., *Now click on the cap*). In utterance-medial position, the word was followed by a prosodic-word boundary, whereas in utterance-final position, the word was followed by a stronger prosodic boundary, namely an utterance boundary. Because the degree of lengthening of the final segments of a word was expected, on the basis of the phonetics literature, to be proportional to the size of the boundary following the word, it was expected that the final segments of the word would be of longer duration when the word occurred in final position than when it occurred in medial position. In line with these predictions, words in utterance-medial position

(which were followed by a prosodic-word boundary, which is a minor prosodic boundary) were of shorter duration than words in utterance-final position (which were followed by an utterance boundary, which is a major prosodic boundary). Experiments 3.1 and 3.2 examined the consequences of this prosodically-conditioned speech variation on the identification of a word and on the degree to which similar-sounding competitor words are considered for recognition.

In these two eye-tracking experiments, listeners carried out spoken instructions to manipulate one of four objects on a computer screen. Displayed along with the target picture were either one (in Experiment 3.1) or two (in Experiment 3.2) pictures that represented competitor words starting with the same sounds as the target word. The ease with which the target word (e.g., *cap*) was identified and the degree to which different types of competitor words (e.g., *cat* or *captain*) were considered for recognition were estimated from participants' eye movements to pictures in the visual display upon hearing the name of the target object.

The results of Chapter 3 revealed two main findings. First, target words in utterance-final position were identified more easily than target words in utterance-medial position. This suggests that the mapping of the speech signal onto lexical representations was facilitated in utterance-final position compared to utterance-medial position. Second, the pattern of lexical activation associated with the processing of a monosyllabic word (which was the main focus of experiments reported in Chapter 3) was affected by the word's position in the sentence. The processing of a monosyllabic word (e.g., *cap*) in utterance-final position, compared to processing of the same word in utterance-medial position, was associated with an increase in lexical activation of monosyllabic competitors (e.g., *cat*) and with a decrease in lexical activation of polysyllabic competitors (e.g., *captain*). This demonstrates that the pattern of competitor activation that is associated with the processing of a spoken word can be affected by that word's position in an utterance. A particular competitor may interfere with the recognition of a spoken word more strongly when the word occurs in one utterance position than when it occurs in another utterance position. The results of Chapter 3 thus converge with the results of Chapter 2 by providing clear evidence for listeners' use of subphonemic information conditioned by prosodic structure in lexical processing.

IMPLICATIONS FOR MODELS OF SPOKEN-WORD RECOGNITION

Taken together, the results reported in this thesis provide strong evidence that listeners exploit speech variation conditioned by prosodic structure to assist the recognition

of spoken words. This is an important finding because it demonstrates that speech variation associated with prosodic structure is relevant for lexical processing. In order for this information to be exploited by listeners, it has to be preserved in the representations that mediate the recognition of spoken words. In order to account for the findings of this thesis, models of spoken-word recognition should therefore incorporate representations that can capture subphonemic information associated with (constituent-level) prosodic structure. One class of models that can be ruled out on the basis of the findings of this dissertation is the class of models that rely only on purely phonemic representations. Information contained in such representations discards any information that is not phonemically contrastive. Phonemic representations can therefore not capture differences in realization between a monosyllabic word and the first syllable of a polysyllabic word that has the monosyllabic word phonemically embedded at its onset.

The experiments reported in this thesis did not directly address the question of how and where in the word-recognition system prosodically-conditioned subphonemic variation is represented. At the very least, such information must be contained in some mental representation of the speech signal. In order for models that do not incorporate prelexical representations (i.e., direct-mapping models) to account successfully for these findings, speech variation that is associated with prosodic structure must be represented in lexical representations. For instance, the lexical representations of monosyllabic words would be characterized by longer durations, while the initial portion of polysyllabic words would be characterized by shorter durations. A spoken sequence with a relatively long duration would thus provide a closer match to the lexical representation of a monosyllabic word than it would to the lexical representation of a polysyllabic word.

In models with prelexical representations, prosodically-conditioned speech variation may be represented in lexical representations, but it need not be. Lexical representations can, for instance, be purely phonemic in nature, as long as representations at the prelexical level are sensitive to prosodically-conditioned speech variation, and as long as these representations can influence the activation of lexical representations. In Chapter 2, a model of spoken-word recognition was proposed that satisfies these constraints. In this model, subphonemic information associated with prosodic structure is used to construct a prosodic representation of the speech signal, in tandem with a phonemic encoding of the speech signal. Aspects of this prosodic representation of the speech input, such as the edges of prosodic constituents equal to or higher than the word, coincide with the location of word boundaries. This information is then

used to assist the evaluation of lexical candidates, for example by favoring lexical hypotheses that are aligned with word boundaries.

The finding that the phonetic manifestation of prosodic structure has an impact on lexical activation raises the question of how such phonetic information is evaluated and processed by the word-recognition system. For instance, durational information in the speech signal is associated with the edges of prosodic constituents, but it also varies as a function of other factors, such as speech rate. The relationship between the duration of segments in the speech signal and the location of word boundaries is therefore probabilistic. This was evident in Experiments 2.2 and 2.3 in Chapter 2, which demonstrated that listeners' preferred lexical interpretation of a spoken sequence (i.e., whether it corresponded to a monosyllabic word or to the onset of a polysyllabic word) was affected by the duration of that sequence. When the sequence was of relatively long duration, it generated more monosyllabic-word interpretations than when it was of relatively short duration. A speech segment or sequence is, however, not intrinsically long or short. In order for a spoken sequence to be interpreted as being of relatively long or short duration thus appears to require an evaluation of the sequence's duration in a larger context, for instance the average duration of segments immediately preceding it.

That spoken-word recognition entails the continuous and parallel evaluation of lexical hypotheses in response to an unfolding speech signal is well established. The simultaneous activation of lexical hypotheses whose sound form is similar to that of the unfolding speech signal can be viewed as a pattern resembling a landscape, with some hypotheses corresponding to more prominent features than others. In many models of spoken-word recognition, the structure of the landscape associated with a particular word is, by and large, fixed. That is, the most prominent features of the landscape always correspond to the same lexical hypotheses: the degree to which each of those features emerges from that landscape and dominates its appearance is thus more or less constant. The main contribution of the present research is that it suggests instead that the structure of the landscape of simultaneously activated lexical hypotheses associated with an unfolding spoken word is established dynamically on the basis of fine-grained, prosodically-conditioned subphonemic information in the speech input, and that the structure of the resulting landscape affects the recognition of that word.

RICHNESS OF PROSODIC CUES

Prosodic structure has many different acoustic manifestations. Although it is clear that the representations that mediate the mapping of the speech signal onto stored lexical knowledge must be sensitive to at least some acoustic correlates of prosodic structure, the fact that prosodic structure is manifested in many different ways in the speech signal warrants a detailed investigation of which types of speech variation associated with prosodic structure are relevant for lexical processing. Different types of prosodically-conditioned speech variation, such as initial and final lengthening, pitch accents, pitch movement and articulatory strengthening, may each be exploited in different ways during spoken-word recognition. This may depend, in part, on the time course with which each of these types of information becomes available in the speech signal. It remains to be seen whether all of the manifestations of prosodic structure influence spoken-word recognition and whether their effects interact.

Furthermore, prosodically-conditioned speech variation is likely to affect lexical processing in different ways, for example by facilitating the mapping of the speech signal onto lexical representations, or by reducing the lexical ambiguity inherently associated with partial spoken input. The research presented in this thesis demonstrated that the interpretation of partial spoken input is affected by the degree to which cues associated with constituent-level prosodic structure enhance differences in realization between similar-sounding words. The exact role of other types of prosodically-conditioned subphonemic variation in lexical processing is a topic of future investigation.

CONCLUSION

The research presented in this thesis provides strong evidence that subphonemic speech variation conditioned by constituent-level prosodic structure affects the recognition of spoken words. Chapter 2 demonstrated that listeners are sensitive to subphonemic, prosodically-conditioned differences between monosyllabic words and the (phonemically identical) first syllable of polysyllabic words. Chapter 3 showed that listeners are also sensitive to differences in the realization of words across different, prosodically-defined positions in an utterance. Spoken-word recognition can therefore not rely exclusively on purely phonemic representations, and models that incorporate such representations need to be revised in order to accommodate the present research. The findings of this thesis also attest to the extraordinary sensitivity of the spoken-word recognition system by demonstrating the relevance for lexical processing of very fine-grained phonetic detail.

SAMENVATTING

Luisteraars zijn zich—en dat is maar goed ook—in het dagelijks leven niet bewust van de ingewikkelde cognitieve processen die hen in staat stellen gesproken taal te begrijpen. Een van de problemen waarmee de luisteraar zich geconfronteerd ziet is dat hij om een gesproken boodschap te begrijpen alle woorden in het spraaksignaal moet herkennen. Het spraaksignaal is tijdelijk van aard, zeer variabel en constant in beweging. Om de woorden in dit signaal te herkennen dient de luisteraar informatie in het spraaksignaal te vergelijken met de representaties van de klankvormen van woorden in zijn geheugen, zogenaamde lexicale representaties.

Een belangrijk kenmerk van gesproken taal is dat het spraaksignaal bestaat uit een vrijwel constante stroom van klanken. Hierin onderscheidt gesproken taal zich van geschreven taal, waarin woorden van elkaar gescheiden zijn door spaties. Een geschreven woord kan bovendien in zijn geheel worden waargenomen, terwijl een gesproken woord slechts bestaat zolang het uitgesproken wordt. Het verwerken van geschreven taal is hierdoor minder ingewikkeld dan het verwerken van gesproken taal; de lezer heeft het over het algemeen een stuk makkelijker dan de luisteraar. Om gesproken taal te begrijpen moet de luisteraar namelijk gelijke tred houden met informatie in het spraaksignaal. Luisteraars doen dit door informatie in het spraaksignaal direct en continu te analyseren en te verwerken. Wanneer een luisteraar de eerste klanken van een woord heeft gehoord, worden in zijn geheugen alle woorden geactiveerd die met die klanken beginnen. Hoort een luisteraar bijvoorbeeld *ha...*, dan worden in zijn geheugen woorden zoals *ham, hak, hamster* en *handel* geactiveerd. Door als het ware bij te houden wat het woord in het spraaksignaal zou kunnen zijn, kan de luisteraar dat woord zeer snel herkennen. Wanneer hij bijvoorbeeld *hams...* heeft gehoord, is er nog maar een woord in zijn geheugen geactiveerd dat met deze klanken begint, namelijk het woord *hamster*. Hierdoor kan een luisteraar een woord vaak herkennen voordat hij het in zijn geheel gehoord heeft.

Het spraaksignaal bevat een overweldigende hoeveelheid akoestische informatie. Een en hetzelfde woord kan op vele verschillende manieren worden uitgesproken en zelfs wanneer een spreker dat woord tien keer achter elkaar uitspreekt, zal het woord nooit precies hetzelfde klinken. Een belangrijke vraag in onderzoek naar gesproken woordherkenning is welke soorten akoestische informatie de luisteraar gebruikt bij

het herkennen van gesproken taal. In dit proefschrift werd onderzocht of en in welke mate akoestische variatie in het spraaksignaal die verband houdt met prosodische structuur van invloed is op het woordherkenningsproces.

Prosodie is een abstracte structuur die de groepering en benadrukking van spraakklanken beïnvloedt. Deze structuur vormt een belangrijke bron van systematische akoestische variatie in het spraaksignaal. Zo is de duur van een spraakklank bijvoorbeeld vaak iets langer aan het eind van een woord dan midden in een woord. Een luisteraar zou deze informatie kunnen gebruiken om het woordherkenningsproces iets efficiënter te laten verlopen. Wanneer hij een spreker *ha*.. hoort zeggen en deze klankreeks aan de lange kant is, zou de luisteraar kunnen veronderstellen dat de spreker een kort woord, zoals *ham*, uitspreekt en niet een lang woord, zoals *hamster*. De duur van spraakklanken en de manier waarop een woord wordt uitgesproken worden echter beïnvloed door tal van andere factoren. Daarom is meestal aangenomen dat de akoestische manifestatie van prosodische structuur in het spraaksignaal hoogstens een marginale invloed kan hebben op het herkennen van gesproken woorden.

In hoofdstuk 2 van dit proefschrift werd de verwerking onderzocht van lettergrepen die uit een identieke reeks van klanken bestaan maar die afkomstig waren uit woorden van verschillende lengte, zoals bijvoorbeeld het woord *ham* en de eerste lettergreep van het woord *hamster*. Deze woorden waren uitgesproken in het midden van een zin, bijvoorbeeld *Ze dacht dat die hamster verdwenen was*. Hoewel de lettergrepen dus bestonden uit dezelfde reeks van klanken, verschilden zij met betrekking tot hun positie in de prosodische structuur van de zin. Het woord *ham* wordt in een dergelijke structuur namelijk gevolgd door een prosodische woordgrens, terwijl dit niet het geval is voor de eerste lettergreep van het woord *hamster*. Op grond hiervan werd verwacht dat er kleine verschillen zouden zijn in de uitspraak van de lettergrepen. Dit bleek inderdaad het geval: de lettergreep *ham* was, gemiddeld genomen, iets langer van duur in het woord *ham* dan in het woord *hamster*. In hoofdstuk 2 werd onderzocht of subtiele verschillen in uitspraak tussen woorden van verschillende lengte die met dezelfde klanken beginnen het woordherkenningsproces zouden kunnen beïnvloeden.

In een reeks van experimenten werden de oogbewegingen van Nederlandse luisteraars geregistreerd terwijl ze keken naar een aantal plaatjes op een computerscherm en tegelijkertijd luisterden naar gesproken zinnen. In de zin werd een van de plaatjes genoemd. De proefpersonen was voorafgaand aan het experiment gevraagd het plaatje dat in de zin genoemd werd te verplaatsen met de muis. De mate waarin de proefpersonen tijdens de presentatie van de zin keken naar een bepaald plaatje werd gebruikt als een indicatie van hun lexicale interpretatie van het spraaksignaal. Verwacht werd dus dat de proefpersonen bijvoorbeeld tijdens de presentatie van de zin *Ze dacht dat*

die hamster verdwenen was niet alleen naar een plaatje van een hamster zouden kijken, maar soms ook heel even naar een plaatje van een ham, omdat de eerste klanken van het woord *hamster* overeenkomen met de eerste klanken van het woord *ham*.

De zinnen die tijdens het experiment werden gebruikt waren op subtiele wijze gemanipuleerd. In de zin *Ze dacht dat die hamster verdwenen was*, was de eerste lettergreep van het woord *hamster* namelijk ofwel vervangen door het woord *ham*, afkomstig uit een andere zin (*Ze dacht dat die ham stukgesneden was*), ofwel door de eerste lettergreep van het woord *hamster*, afkomstig uit een andere opname van de zin *Ze dacht dat die hamster verdwenen wa*s. Dit leverde twee versies van een zin op die subtiel verschilden met betrekking tot de uitspraak van de eerste lettergreep van het woord *hamster*. Hoewel proefpersonen bij het luisteren naar beide zinnen dachten dat zij het woord *hamster* hoorden, was de eerste lettergreep van dit woord in een van de zinnen oorspronkelijk uitgesproken als het woord *ham*.

In het eerste experiment van hoofdstuk 2 werd aangetoond dat de lexicale interpretatie van de eerste lettergreep van het woord *hamster* beïnvloed werd door subtiele variatie in het spraaksignaal. Proefpersonen keken meer naar het plaatje van een ham wanneer de eerste lettergreep van het woord *hamster* oorspronkelijk was uitgesproken als het woord *ham* dan wanneer deze lettergreep afkomstig was uit het woord *hamster*. Dit effect werd echter alleen gevonden in experiment 2.1A, waarin het woord *ham* afkomstig was uit een zinscontext waarin de duur van dit woord 20 ms verschilde van de duur van de eerste lettergreep van het woord *hamster*. In experiment 2.1B, waarin het woord *ham* afkomstig was uit een zinscontext waarin de duur van dit woord slechts 15 ms verschilde van de duur van het woord *hamster*, werd weliswaar een vergelijkbaar effect gevonden, maar dit effect was zwakker en niet statistisch significant. De resultaten van het eerste experiment van hoofdstuk 2 lieten dus zowel zien dat subtiele variatie in het spraaksignaal die verband houdt met prosodische structuur de verwerking van gesproken woorden kan beïnvloeden (experiment 2.1A), terwijl de resultaten ook aangeven dat dergelijke akoestische informatie, zelfs wanneer zij aanwezig is in het spraaksignaal, niet zonder meer een invloed heeft op het woordherkenningsproces (experiment 2.1B).

Op basis van deze resultaten werd de hypothese ontwikkeld dat de mate waarin luisteraars het begin van het woord *hamster* tijdelijk interpreteerden als het woord *ham* bepaald werd door de mate waarin de uitspraak van de lettergreep *ham* (als het woord *ham* of als eerste lettergreep van het woord *hamster*) beïnvloed was door de prosodische structuur van de zin. In een analyse waarin alle woorden die in het experiment gebruikt waren betrokken werden, bleek dat er een sterk verband was tussen de mate waarin de duur van het kortere woord en de eerste lettergreep van het langere

woord dat met dezelfde klanken begon van elkaar verschilden en het verschil in de mate waarin luisteraars keken naar het plaatje dat het korte woord voorstelde. In twee vervolgexperimenten werd de invloed van de duur van de lettergrepen (d.w.z., het woord *ham* en de eerste lettergreep van het woord *hamster*) op het woordherkennigsproces verder onderzocht. Indien de duur van de sequentie verband houdt met de mate waarin een woordgrens aanwezig is in de prosodische structuur en indien deze informatie een invloed heeft op het woordherkenningsproces, werd verwacht dat de lexicale interpretatie van het spraaksignaal voorspeld kon worden op grond van de duur van de lettergrepen.

Voor het eerste experiment van hoofdstuk 1 waren meerdere exemplaren van iedere zin opgenomen. Het tweede experiment was vrijwel identiek aan dit experiment, met het verschil dat voor dit experiment zinnen uit de oorspronkelijke opname werden geselecteerd op grond van de duur van de lettergrepen. Op die manier werden, net als in het eerste experiment, twee varianten gemaakt van bijvoorbeeld de zin *Ze dacht dat die hamster verdwenen was.* In de ene zin was de eerste lettergreep van het woord *hamster* oorspronkelijk uitgesproken als het woord *ham* en in de andere zin was de eerste lettergreep van het woord *hamster* oorspronkelijk uitgesproken als het begin van het woord *hamster*. Echter, in het tweede experiment werden deze lettergrepen zodanig geselecteerd dat het verschil in duur tussen de lettergrepen minimaal was. In overeenstemming met de verwachtingen bleek dat de mate waarin proefpersonen keken naar het plaatje van de ham niet verschilde voor de twee varianten van de gesproken zin.

In het derde experiment van hoofdstuk 1 werden zinnen zodanig samengesteld dat de eerste lettergreep van het woord *hamster* van kortere duur was wanneer de lettergreep oorspronkelijk was uitgesproken als het woord *ham* dan wanneer de lettergreep afkomstig was van het begin van het woord *hamster*. (Dit verschil in duur is precies tegengesteld aan het patroon dat gevonden wordt in gesproken Nederlands, waarin de duur van het woord *ham* meestal iets langer is dan de duur van de eerste lettergreep van het woord *hamster*.) De verschillen in uitspraak tussen de eerste lettergreep van het woord *hamster* in de twee varianten van de zin bleken een invloed te hebben op de lexicale verwerking van het woord *hamster*. Proefpersonen keken ditmaal meer naar het plaatje van een ham wanneer de eerste lettergreep van het woord *hamster* afkomstig was uit een andere opname van het woord *hamster* dan wanneer deze lettergreep oorspronkelijk was uitgesproken als het woord *ham*.

De experimenten in het eerste hoofdstuk van dit proefschrift tonen aan dat luisteraars een gesproken sequentie die in de prosodische structuur wordt gevolgd door een woordgrens (bijvoorbeeld het woord *ham*) kunnen onderscheiden van een identieke

klankreeks die het begin vormt van een langer woord (bijvoorbeeld de eerste lettergreep van het woord *hamster*). De experimenten tonen echter ook aan dat het woord *ham* en de eerste lettergreep van het woord *hamster* niet altijd verschillen in duur. De prosodische structuur van een zin lijkt dus een invloed te hebben op de verwerking van gesproken woorden in zoverre dat wanneer deze structuur zich akoestisch manifesteert in de vorm van variatie in duur van klanken, deze informatie de lexicale verwerking van het spraaksignaal door luisteraars beïnvloedt.

In hoofdstuk 3 werd de invloed van de positie van een woord in een gesproken zin op het lexicale verwerkingsproces onderzocht. De akoestische realisatie van een gesproken woord wordt nameljk beïnvloed door de positie van het woord in de prosodische structuur van een uiting. Daarom zou de positie van een woord in een zin een invloed kunnen hebben op de lexicale verwerking van dat woord. In hoofdstuk 3 werd onderzocht of de positie van een woord in een zin invloed heeft op het gemak waarmee het woord herkend kan worden en de mate waarin de lexicale verwerking van het woord leidt tot gelijktijdige activatie van verschillende soorten woorden die met dezelfde klanken beginnen.

In de experimenten in hoofdstuk 3, die werden uitgevoerd in de Verenigde Staten, werd de verwerking van een woord in het midden van een zin vergeleken met de verwerking van datzelfde woord aan het eind van een zin. In het midden van een zin wordt een woord gevolgd door een prosodische woordgrens (bijvoorbeeld in de Engelse zin *Put the cap next to the square*; zet de pet naast het vierkant), terwijl een woord aan het eind van een zin gevolgd wordt door een sterkere prosodische grens, namelijk een zinsgrens (bijvoorbeeld in de Engelse zin *Now click on the cap*; klik nu op de pet). Klanken aan het eind van een woord worden verlengd naarmate de prosodische grens die op het woord volgt sterker is. Daarom werd verwacht dat de klanken aan het eind van een woord van langere duur zouden zijn voor een woord aan het eind van een zin dan voor datzelfde woord midden in een zin. Deze verwachting werd bevestigd. In twee experimenten werd onderzocht wat voor gevolgen deze door prosodische structuur bepaalde variatie in de realisatie van het woord had op de herkenning van het woord en op de mate waarin andere woorden die met dezelfde klanken beginnen het woordherkenningsproces beïnvloeden.

In twee oogbewegingsexperimenten voerden Amerikaanse luisteraars gesproken instructies uit om een plaatje op een computerscherm te verplaatsen (*Put the cap next to the square*) of om op dat plaatje te klikken (*Now click on the cap*). Het woord dat in de zin genoemd werd bestond altijd uit een lettergreep (bijvoorbeeld het Engelse woord *cap*, pet). Op het scherm stonden vier plaatjes waaronder één (in Experiment 3.1) of twee (Experiment 3.2) plaatjes waarvan de naam met dezelfde klanken als het

woord in de zin begon. Onderzocht werd of de positie van het woord in de zin een invloed had op het gemak waarmee dat woord herkend werd en op de mate waarin woorden van verschillende lengte die met dezelfde klanken beginnen (bijvoorbeeld de Engelse woorden *cat*, poes, en *captain*, kapitein) geactiveerd werden en van invloed waren op het woordherkenningsproces.

De experimenten in hoofdstuk 3 leverden twee belangrijke resultaten op. Allereerst werd aangetoond dat woorden aan het eind van een zin gemakkelijker herkend worden dan woorden in het midden van een zin. Dit suggereert dat het activeren van lexicale representaties in het geheugen op grond van informatie in het spraaksignaal gemakkelijker verloopt aan het eind van een zin dan in het midden van een zin. Een tweede bevinding was dat de positie van het woord in de zin (bijvoorbeeld *cap*) een verschillende invloed had op de mate waarin woorden van verschillende lengte (bijvoorbeeld *cat* en *captain*) geactiveerd werden tijdens het woordherkenningsproces. Het woord *cat*, dat net als het woord *cap* uit één lettergreep bestaat en in de prosodische structuur van een uiting altijd wordt gevolgd door een woordgrens, werd sterker geactiveerd aan het eind van een zin dan in het midden van een zin. Daarentegen werd het woord *captain* juist sterker geactiveerd in het midden van een zin dan aan het eind van een zin. De mate waarin tijdens de herkenning van een woord andere woorden die met dezelfde klanken beginnen in aanmerking worden genomen kan dus beïnvloed worden door de positie van het woord in de prosodische structuur van een uiting. Deze andere woorden hebben een invloed op het woordherkenningsproces, maar de mate waarin zij dat proces beïnvloeden blijkt afhankelijk te zijn van de positie in de zin waarin het woord dat herkend wordt zich bevindt.

De bevindingen in dit proefschrift tonen aan dat informatie in het spraaksignaal die verband houdt met de realisatie van bepaalde aspecten van de prosodische structuur van een uiting door luisteraars gebruikt kan worden bij het herkennen van gesproken woorden.

CURRICULUM VITAE

Anne Pier Salverda werd geboren in Wageningen op 26 april 1975. Na het behalen van zijn VWO-diploma aan het Peelland College Deurne studeerde hij Psychologie aan de Katholieke Universiteit Nijmegen. Hij behaalde zijn doctoraalexamen Psychologische Functieleer in 1998 en werkte daarna als onderzoeksassistent aan het Max Planck Instituut (MPI) voor Psycholinguïstiek in Nijmegen. In November 1999 werd hem een stipendium toegekend door het Max-Planck-Gesellschaft om promotieonderzoek te doen binnen de Comprehension Group van het MPI. In 2002 werkte hij op uitnodiging van professor Michael Tanenhaus een jaar in het Department of Brain and Cognitive Sciences aan de University of Rochester (VS). Vanaf oktober 2003 was hij als research fellow verbonden aan het Department of Psychology van de University of York (GB). Sinds april 2005 werkt hij als postdoctoral fellow in het Department of Brain and Cognitive Sciences aan de University of Rochester.

MPI SERIES IN PSYCHOLINGUISTICS

1. The electrophysiology of speaking: Investigations on the time course of semantic, syntactic, and phonological processing
   Miranda van Turennout

2. The role of the syllable in speech production: Evidence from lexical statistics, metalinguistics, masked priming, and electromagnetic midsaggital articulography
   Niels O. Schiller

3. Lexical access in the production of ellipsis and pronouns
   Bernadette M. Schmitt

4. The open-/closed-class distinction in spoken-word recognition
   Alette Haveman

5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach
   Kay Behnke

6. Gesture and speech production
   Jan-Peter de Ruiter

7. Comparative intonational phonology: English and German
   Esther Grabe

8. Finiteness in adult and child German
   Ingeborg Lasser

9. Language input for word discovery
   Joost van de Weijer

10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe
    James Essegbey

11. Producing past and plural inflections
    Dirk Janssen

12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea
    Anna Margetts

13. From speech to words
    Arie van der Lugt

14. Simple and complex verbs in Jaminjung: A study of event categorisation in an Australian language
    Eva Schultze-Berndt

15. Interpreting indefinites: An experimental study of children's language comprehension
    Irene Krämer

16. Language-specific listening: The case of phonetic sequences
    Andrea Weber

17. Moving eyes and naming objects
Femke van der Meulen

18. Analogy in morphology: The selection of linking elements in Dutch compounds
Andrea Krott

19. Morphology in speech comprehension
Kerstin Mauth

20. Morphological families in the mental lexicon
Nivja H. de Jong

21. Fixed expressions and the production of idioms
Simone A. Sprenger

22. The grammatical coding of postural semantics in Goemai (a West Chadic language of Nigeria)
Birgit Hellwig

23. Paradigmatic structures in morphological processing: Computational and cross-linguistic experimental studies
Fermín Moscoso del Prado Martín

24. Contextual influences on spoken-word processing: An electrophysiological approach
Daniëlle van den Brink

25. Perceptual relevance of prevoicing in Dutch
Petra M. van Alphen

26. Syllables in speech production: Effects of syllable preparation and syllable frequency
Joana Cholin

27. Producing complex spoken numerals for time and space
Marjolein Meeuwissen

28. Morphology in auditory lexical processing: Sensitivity to fine phonetic detail and insensitivity to suffix reduction
Rachèl J. J. K. Kemps

29. At the same time...: The expression of simultaneity in learner varieties
Barbara Schmiedtová

30. A grammar of Jalonke argument structure
Friederike Lüpke

31. Agrammatic comprehension: An electrophysiological approach
Marlies Wassenaar

32. The structure and use of shape-based noun classes in Miraña (North West Amazon)
Frank Seifart

33. Prosodically-conditioned detail in the recognition of spoken words
Anne Pier Salverda