FROM SPEECH TO WORDS

Design: Linda van den Akker, Inge Doehring
Illustration: Oskar Schlemmer, Abstract figure © Kröller-Müller Museum, Otterlo
Printed and bound by: Ponsen & Looijen bv, Wageningen

FROM SPEECH TO WORDS

een wetenschappelijke proeve
op het gebied van de Sociale Wetenschappen

Proefschrift
ter verkrijging van de graad van doctor
aan de Katholieke Universiteit Nijmegen,
volgens besluit van het College van Decanen
in het openbaar te verdedigen
op maandag 29 november
des namiddags om 1.30 uur precies

door
Arie Huibrecht van der Lugt
geboren op 20 juli 1968 te Ede

Promotor: Prof. dr. A. Cutler
Co-promotor: Dr. J. M. McQueen

Manuscriptcommissie:
Prof. dr. H. Schriefers
Prof. dr. U. Frauenfelder (Université Genève)
Prof. dr. P. Zwitserlood (Westfälische Wilhelms-Universität, Münster)

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

Voor Sum & Mok
en
Roosje

VOORWOORD

Summer time and the living is easy...belooft Billie Holiday me aan het einde van iedere nachtdienst. Even onwrikbaar is het vertrouwen van Anne Cutler in een goede afloop...*write the damn thing!*...en altijd als me dat vertrouwen even te veel wordt, is er James McQueen die behalve een voortreffelijke onderzoeker, onderwijzer en kok ook een perfecte begeleider is. Anne en James bedank ik voor meer dan vijf jaar fijne samenwerking en gezelligheid binnen en buiten het instituut. Gezellig was het ook door het volleybal op dinsdag, de koffie om tien uur en later de etentjes met de dames (met name Sylvia Aal, Rian Zondervan en Edith Sjoerdsma), de Maxkrant, de tijd in Trieste met Niels Schiller, dansen in de Extase, het roken voor de deur met andere bannelingen en de collega-PhD-studenten en assistenten. In het bijzonder wil ik Dirk 'Bep' Janssen bedanken voor het gezellig samen regelneven (uiteindelijk heeft hij zelfs een baan voor me geregeld) en mijn kamergenoten Nicole Cooper en Petra van Alphen voor het veelvuldig opfleuren van mijn humeur.

Dank gaat ook uit naar alle ondersteunende mensen van het Max Planck: Hans Gerrits, de mensen van de receptie (Paul, Trix, Mariska en Leontine), Rian Zondervan, de administratie (met name Norbert 'achtvijftig' Nielen), de bibliothecaresses en de technische groep. Rinske Goettsch en Joost van de Weijer bedank ik voor het uitspreken van alle *deemdeeps*, *suupdeuven* en *luutvoems*. Verder wil ik een aantal mensen bedanken voor hun commentaar op delen van het proefschrift: Martin Cooper, Nicole Cooper, Peter Jucszyk, John Kingston, Laura Kroes en Dennis Norris. Bernadette 'the babe' Schmitt bedank ik voor haar mooie plaatjes.

De collega's van JB en Krulliemullie bedank ik voor de goede sfeer, die er mede toe heeft bijgedragen dat ik niet gierend gek ben geworden in de laatste manische maanden. Geweldig waren ook de meiden van de subclub met hun steunbetuigingen voor mijn schrijvende alter-ego Bob. Mijn vrienden en vriendinnen bedank ik voor al het geluk dat ze me geven. Mijn broer Herbert en Niels Schiller bedank ik in het bijzonder omdat ze bij de verdediging mijn paranimfen willen zijn. ...*But until that morning, there's nothing gonna harm you*...Altijd was er al het warme nest van mijn ouders om op terug te vallen, sinds een jaar of zeven komt daar de liefde van Roosje bij.

TABLE OF CONTENTS

Chapter 1 *Introduction* 1

Chapter 2 *Relative positional probabilities of consonant clusters as a cue for segmentation: a first experiment* 13

INTRODUCTION

CHAPTER 1

How do we get from speech to words, from signal to meaning? The study of the comprehension of spoken language is a relatively new field within the domain of psycholinguistics. For a long time research focused on the visual modality and most theories developed to explain visual language processing were thought to apply to the recognition of speech just as well. But speech processing is different. Speech is, for instance, inherently sequential: it is spread over time, whereas written language can be processed in a more parallel manner, at least for each saccade within a word. In the last thirty years the study of speech comprehension as a special and complex language process has become increasingly important.

By definition, speech recognition can be divided into four levels of processing: general sensory processing, prelexical speech processing, lexical processing and sentential processing. General sensory processing is the acoustical processing in the human auditory system that is not specific to speech. Prelexical processing refers to a number of processes that are necessary before the spoken words can be mapped onto stored lexical knowledge. Lexical processing is the mapping of the perceived segments onto stored lexical knowledge; that is, candidate words that fit the input are selected and information about them, such as their meaning, phonology, orthography, etcetera, becomes available. The distinction between prelexical and lexical processing is also known as the difference between activation and selection or between lexical access versus retrieval. In sentential processing words are integrated into representations of complete sentences or even larger stretches of speech. Although it is very difficult to disentangle these different levels of processing completely, this thesis will be mainly concerned with the prelexical processing of speech.

PRELEXICAL PROCESSING OF SPEECH

Prelexical processing is best defined by its ultimate goal: the recognition of words no matter in what context they appear. Prelexical processing prepares the incoming stream of sounds in such a way that we are able to understand what is said. To do this the two basic problems of spoken word recognition need to be solved: the problem of segmentation and the problem of variation.

The first problem prelexical processing has to solve is where the word boundaries are in the continuous stream of speech: the segmentation problem. In written language, spaces unambiguously indicate word boundaries; in speech such a clear demarcation does not exist. The ease with which we perceive individual words in our own language is misleading. This becomes painfully clear when you try to understand a language that you don't speak. For example, when a monolingual Slovenian visitor at the Kröller-Müller museum has lost her group. The problem of segmentation is the topic of the first part of this thesis (Chapters 2 and 3).

The second problem prelexical processing needs to solve is the problem of variation. There is huge variation in the realisation of speech sounds due to many factors: the idiosyncrasies of each individual speaker's voice, the sex of the speaker, speech rate, dialect, noise, coarticulation, etcetera. Nevertheless, human speech perception is remarkably stable. Most of the early studies on speech perception were concerned with how individual speech sounds were recovered from the variable speech signal (for an overview see Repp & Liberman, 1987). More recently, an increasing amount of research focuses on how the auditory signal is mapped onto stored lexical knowledge. Some propose that the mapping is done directly in spite of the variation (Klatt, 1989), but this seems implausible. The variability of the speech signal requires abstraction away from the actual input to a somewhat normalised intermediate level of representation, but what is this 'unit of perception' ? Is it the acoustic-phonetic feature (Eimas & Corbit, 1973), is it the articulatory gesture (Liberman & Mattingley, 1985), is it the syllable (Mehler, Dommergues, Frauenfelder & Segui, 1981) or is it the phoneme (Fowler, 1984)? There seems to be evidence for all units of perception and perhaps it is indeed the case that listeners use different units of perception, depending on the (experimental) situation (Pisoni & Luce, 1987). Besides these representational issues there are some processes of normalisation that have been shown to take place: rate normalisation (Miller & Liberman, 1979) and speaker normalisation (Mullenix, Pisoni & Martin, 1989; Mullenix & Pisoni, 1990).

Another important issue in dealing with the problem of variation is the tolerance of the word recognition system to mismatching information. In order to adjust flexibly to changing contexts one would expect some degree of flexibility in the process of word recognition with respect to imperfect realisations of words. The classical example was introduced by Norris (1982): spoken word recognition should be able to deal with a mispronunciation like *shigarette*. The sensitivity (or insensitivity) of spoken word recognition with respect to misleading information is the focus of the second part of this thesis (Chapters 4 and 5).

FUNCTIONAL MODELS OF SPOKEN WORD RECOGNITION

In the domain of psycholinguistics several theories have been developed that try to account for the strategies and mechanisms people use for the recognition of spoken words. Here, the three most influential models will be discussed briefly with respect to how they deal with segmentation and mismatching information.

Anticipation in the COHORT model

The anticipation strategy of the original COHORT model (Marslen-Wilson & Welsh, 1978; Marslen-Wilson & Tyler, 1980) is a strictly sequential process[1]: all words beginning in the same way are accessed and constitute a 'cohort' of candidate words, mismatching words drop out as soon as mismatching information comes in, and the word is recognised when it becomes unique. After recognition of one word, the onset for the next cohort can be anticipated to start at the end of that word. The next cohort can therefore be accessed to enable recognition of the following word.

In the COHORT model, any mismatch completely prevents words from being activated as candidates.

There is a growing body of evidence in the literature that shows that the COHORT view and its anticipation strategy is not an adequate mechanism to explain word recognition. The activation of misaligned words and suffixed words like *christmassy* (Shillcock, 1990) , the multitude of initial embeddings (McQueen, Cutler, Briscoe & Norris, 1995) and the importance of information following a word's offset (Bard, Shillcock & Altmann, 1988) all pose insoluble problems for a strictly sequential model of word segmentation. Other major deficits of this theory are the absence of a good backtracking mechanism which can operate when mistakes are made, and the inability to run simulations

---

[1] *A similar model was developed by Cole and Jakimik (1978, 1980).*

because the strictly sequential COHORT segmentation strategy has never been implemented. Although the anticipation strategy is not useful in those cases where a word does not become unique until or even after its offset (cf. Luce, 1986), it could be a good strategy in those cases where a word becomes unique before its offset.

Competition and segmentation: TRACE and SHORTLIST

A major advantage of TRACE (Elman & McClelland, 1988; McClelland 1979, 1987, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981) and SHORTLIST (Norris, 1994; Norris, McQueen, Cutler & Butterfield, 1997) over the COHORT model is that both of these models have been computationally implemented. Simulations have put their actual performance to the test and both models have been used to simulate successfully a large body of experimental data. The core mechanism of TRACE and SHORTLIST is competition. In response to the input, candidate words become activated and compete with one another for recognition. Lexical competition is a powerful mechanism to explain the speed of word recognition; it does not rely too much on information about the locus of word onsets. In other words, Cohort-like behaviour may be seen for early-unique words, but a word that doesn't become unique before its offset poses no problem to the system. A lot of indirect support for competition is provided by studies that show evidence for multiple activation of candidate words (Marslen-Wilson 1987, 1990; Shillcock, 1990; Zwitserlood, 1989; Gow & Gordon, 1995). Phonological priming studies have shown inhibitory effects which probably reflect competition (Goldinger, Luce & Pisoni, 1989; Goldinger, Luce, Pisoni & Marcario, 1992; Slowiaczek & Hamburger, 1992; Radeau, Morais & Dewier, 1989). The most direct evidence for competition comes from a study by McQueen, Norris and Cutler (1994). In a series of three experiments they use a word-spotting task to investigate competition effects. Participants found it easier to detect the English word *mess* when it was embedded in *nemess* which is not the beginning of an existing English word, than when it was embedded in the nonsense string *domess* which is the beginning of the word *domestic*. Listeners also found it easier to detect *sack* in *sacrif*, the onset of *sacrifice*, than in *sackrek*. These results thus show competition effects between both misaligned (*mess* and *domestic*) and aligned (*sack* and *sacrifice*) candidates.

The TRACE model

In the TRACE model, units at three processing levels (representing features, phonemes and words) compete via intra-level inhibition. They also support each other through inter-level facilitation, both bottom-up and top-down. At the lexical level all of the words in the lexicon compete for recognition. Recognition is based on a simple rule that selects the candidate with the highest activation over a certain threshold.

In TRACE, mismatching information does not directly affect lexical access. A direct use of mismatching information could have been modelled by inhibitory phoneme-to-word connections (as was also discussed by McClelland and Elman, 1986, p. 55-57), but they decided against it, arguing that word-to-word inhibition would have the same effect. However, this is only true if the mismatching input provides information in favour of another word, rather than a nonword. If the mismatching information supports a nonword, then in cases where the input shares no features with the relevant phoneme of the possible word, the mismatching information will be treated similarly to the absence of matching information by the model. If the mismatching input shares one or more features (note that single feature mismatches are in fact matches on several other features), mismatching inputs will not block lexical access entirely.

The most important limitation of this model is the implausibility of the architecture: because the full lexicon is involved in competition, an enormous number of connections are needed to deal with the time-invariance problem, that is, so that the model is able to recognise words no matter when they begin. Even when TRACE is only used to model the recognition of isolated words, TRACE's architecture puts severe constraints on the size of the lexicon that can be used for simulations (see also Norris, 1994). Furthermore, there are a number of experimental findings which are problematic for TRACE. Most of those are related to the top-down lexical feedback effects that are predicted by the model (Cutler, Mehler, Norris & Segui, 1986; Frauenfelder, Segui & Dijkstra, 1990; McQueen, 1991; Wurm & Samuel, 1997). For instance, in a phoneme monitoring experiment in French, Frauenfelder et al. found that it is equally difficult to detect [t] in the nonword *vocabutaire*, which is very similar to the real word *vocabulaire*, and in *socabutaire*, which is less similar to any existing word in French. This finding is inconsistent with the predictions of the TRACE model. The TRACE model predicts an inhibitory lexical effect on nonwords that are similar to real words; the activation of *vocabulaire* at the lexical level should hinder the activation of the phoneme [t] at the phoneme level via the lexical feedback to the phoneme [1]. However, recent simulations (Norris, McQueen & Cutler, in press) have shown that a version of the TRACE model can be induced to simulate the results by Frauenfelder et al. Nevertheless, other experimental evidence still argues against the TRACE model. For instance, results on compensation for coarticulation (Pitt & McQueen, 1998) and subcategorical mismatch effects (Marslen-Wilson & Warren, 1994; McQueen, Norris & Cutler, in press) challenge TRACE.

The SHORTLIST model

The recognition and segmentation of words in SHORTLIST is also achieved by competition, as in TRACE, but in this case competition only operates at the lexical level. First, a set of candidate words is accessed: words with a large enough bottom-up activation, that is, those with a high enough degree of fit with the input, become members of the shortlist. The shortlist is then wired into a small interactive-activation network, in which the candidate words compete with each other via inhibitory connections. Because only those lexical candidates which sufficiently match the input can enter the competition process, SHORTLIST can use a realistically-sized lexicon of over 20,000 words.

In SHORTLIST, mismatching segments in the input lower the bottom-up support for lexical candidates, making it unlikely that mismatching candidates will enter the shortlist. They will therefore probably not actively compete for recognition. In the current version of the model effects of mismatch can be modelled by using mid-class phoneme transcriptions for inputs that should not act to block lexical access completely.

Over the past few years the SHORTLIST model has been refined in order to account for several experimental findings that suggest that, besides competition between activated words, more explicit segmentation procedures are used by the listeners as well. In the following sections these experimental findings and the adjustments that were made to enable the SHORTLIST model to account for them are described.

Metrical cues to word boundaries

Listeners use the dominant rhythmic structure of their native language to help locate word boundaries. English and Dutch are so-called stress-timed languages: the rhythm is based on the alternation of strong and weak syllables. (Strong syllables are syllables that contain full vowels, weak syllables are syllables that contain a reduced vowel, usually a schwa.) A good segmentation strategy for stress-timed languages is to segment the input at each strong syllable (Cutler & Norris, 1988 for English; Vroomen & de Gelder, 1995 for Dutch). Confirmation of a metrical strategy of segmentation has been obtained in a number of cross-linguistic studies that investigated the possible use of other rhythmic structures in segmentation: in syllable-timed French (Mehler, Dommergues, Frauenfelder & Segui, 1981; Cutler, Mehler, Norris & Segui, 1986) and in mora-timed Japanese (Otake, Hatano, Cutler & Mehler, 1993). The relation of metrical effects to competition processes has been investigated by Norris, McQueen and Cutler (1995) in a series of word-spotting experiments in English. Their study shows that metrical effects cannot be seen as artifacts of lexical competition: people really are sensitive to the metrical structure of an utterance and they use that information to facilitate the recognition of words in continuous speech. A similar result was found for the role of metrical structure in the segmentation of Dutch (Vroomen, van Zon & de Gelder, 1996).

Phonotactic strategies

If people perceive a combination of sounds that is not allowed within a word in their language they can assume a boundary between those sounds. Because different languages allow different combinations of sounds, the phonotactic information itself is language-dependent, analogous to the crosslinguistic differences in rhythmic structure used in segmentation. McQueen (1998) used a word-spotting task to investigate the role of syllable boundaries which are mandatory on phonotactic grounds in Dutch. It was found that detection of the Dutch word *rok* (skirt) is much harder in *fie-drok*, where the target is misaligned with the syllable boundary of the bisyllabic nonword (voiced stops like [d] must be syllable-initial in Dutch), than in *fiem-rok*, where the target is aligned with the syllable boundary forced by the phonotactic rule that [mr] is an illegal onset or coda cluster in Dutch. In a number of word-spotting experiments, it has been found that vowel harmony provides a useful phonotactic cue in Finnish (Suomi, McQueen & Cutler, 1997; Vroomen,Tuomainen & de Gelder, 1998). In Finnish, all the vowels in a word must be of the same harmony class. Therefore, if two syllables contain a vowel belonging to a different class, this signals a word boundary between them. For example, listeners found it easier to detect *hymy* (smile) in *puhymy* than in *pyhymy*.

The Possible Word Constraint

Possible word parsing refers to people's tendency to rule out impossible segmentations: parsings that leave a residue that cannot be a possible word, for instance a single consonant, are considered much less likely than parsings that leave a bit of speech that could be a word. This has been called the Possible Word Constraint (PWC; Norris et al., 1997). Experimental data from word spotting tasks (Norris et al., 1997) show that it is more difficult to recognize words embedded in following or preceding nonsyllabic contexts than in syllabic contexts. For example, it is harder to detect the word *apple* in the nonsense string *fapple* than in the string *vuffapple*. Here silence is used to postulate a clear boundary before the [f] of *fapple* and the [v] of *vuffapple*. In the case of *fapple*, all parses that leave only syllabic sections of speech are segmentations of the nonsense string that do not correspond to the target *apple* (e.g., *fap-ple*), making the target harder to detect. In the case of *vuffapple* one of the possible word parses leaves a segment that corresponds to the target, facilitating the detection of that target. The PWC is implemented in the most recent version of SHORTLIST (Norris et al., 1997). SHORTLIST postulates boundaries via candidate words, or they can be given in the input: cues provided by silence, phonotactics and metrical structure. These boundaries are used by SHORTLIST to calculate the bottom-up support for candidate words: words that violate the PWC with respect to the tentative boundaries are penalised. The PWC thus provides a unified account of word recognition, where the competition process is modulated by the presence of multiple cues to the location of word boundaries in the speech stream in a principled way, that is, with respect to the possible word constraint

METHODOLOGIES

Several experimental methods have been used to investigate how human listeners deal with segmentation and mismatching information. The two paradigms that were used in the present study are illustrated below.

Word spotting

In word spotting (Cutler & Norris, 1988) subjects are asked to listen for existing words in a nonsense context (see also Figure 1-1). Participants hear a list of nonsense bisyllables. On most of the trials no word will be present in the auditory stimulus, for instance *publi*, *deemdeep*, *luutvoem*; in these cases listeners do not need to react. Listeners do have to respond, however, if a real word is embedded somewhere in the auditory stimulus, for example, if they hear *boomdouf*, *choles*, or *hambur* since *boom* (tree), *les* (lesson), and *ham* (ham) are existing Dutch words. Subjects' performance is measured by both their reaction times and error rate. Word spotting was used in experiments reported in Chapters 2, 3, 4 and 5.

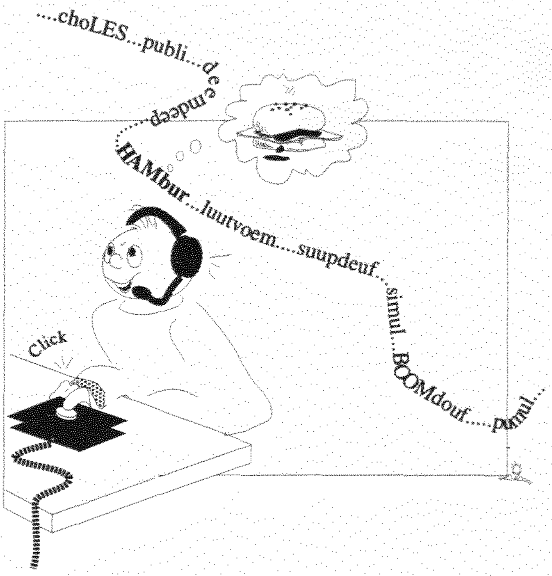

Figure 1-1 *The word spotting task*

Cross-modal identity priming

In identity priming (Zwitserlood, 1996) subjects are asked to listen to a list of auditory stimuli (see also Figure 1-2). At some point in time relative to the auditory stimulus (i.e. the Stimulus Onset Asynchrony) a letter string is presented on a computer screen and the subject is asked to decide whether the letter string forms an existing word or not, by pressing the appropriate response button. On a proportion of the trials participants are presented with a visual stimulus that is related in form to the auditory prime that they have just heard, for instance, they make a lexical decision to the letter string HAM after hearing *hambur*. Performance on these related trials is measured by reaction times and error rates relative to the performance on unrelated trials, where there is no formal relation between auditory prime and visual target, for instance, making a lexical decision to the letter string HAM after hearing *rangcuun*. Cross-modal identity priming tasks were employed in experiments reported in Chapters 4 and 5.

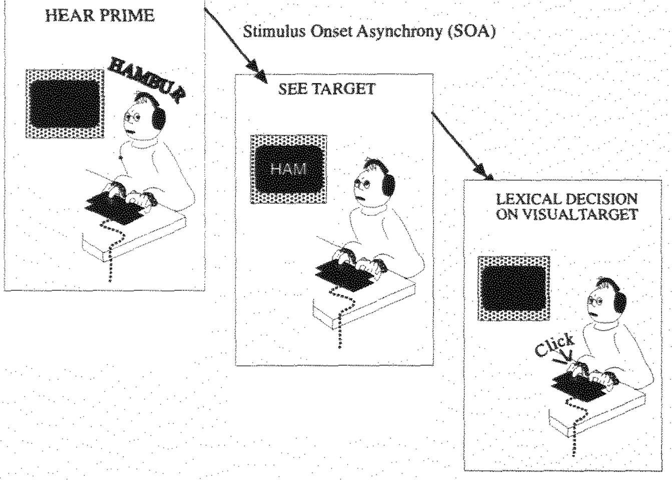

Figure 1-2 *The identity priming task*

STRUCTURE OF THE THESIS

The thesis consists of two parts, corresponding to the two basic problems outline above: the segmentation problem and the problem of speech variation. In the first part (Chapters 2 and 3), eight experiments are reported that investigated the use of sequential probability information in segmentation. In Chapter 2, the first experiment is reported. In this experiment, I investigated the possible use as segmentation cues of the relative positional sequential probabilities of six consonant clusters which can appear both at the beginnings and at the ends of words. In Chapter 3, a series of seven experiments is reported in which the exploitation of absolute positional probability of Consonant-Vowel (CV) and Vowel-Consonant (VC) sequences as segmentation cues was investigated.

In the second part of the thesis (Chapters 4 and 5), ten experiments are described that are concerned with the problem of spoken language variation, specifically, with the role of mismatching information in spoken word recognition. In Chapter 4, four experiments are presented which address the tolerance of the word recognition system with respect to word-initial mismatches. In Chapter 5, six experiments are discussed which focus on mismatches at later positions in a word.

In Chapter 6, the main results of Chapters 2 to 5 are highlighted and tied together. The experimental results are discussed with respect to their implications for theories of word recognition.

RELATIVE POSITIONAL PROBABILITIES OF CONSONANT CLUSTERS AS A CUE FOR SEGMENTATION: A FIRST EXPERIMENT

CHAPTER 2

(slightly adapted version of article published In G. Kokkinakis, N. Fakotakis & E. Dermatas (Eds.), *Eurospeech '97 Proceedings: ESCA 5th European Conference on Speech Communication and Technology*, *4* (2151-2154). 1997, University of Patras, Greece: WCL.)

A. H. van der Lugt

ABSTRACT

A large amount of psycholinguistic research, phonetic research and research in speech technology has been dedicated to the problem of segmentation: how is speech segmented into words? The work reported here extends earlier findings by McQueen (1998), showing that phonotactics are used by listeners as a cue to the location of word boundaries. The present investigation addresses the question of whether people can also use less extreme sequential probabilities as a segmentation cue. Hearing a combination of sounds that often occurs at the end of a word or syllable may facilitate recognition of a following word; hearing a combination of sounds that occurs often at the beginning of a word or syllable may facilitate recognition of a preceding word. In a word-spotting task six consonant clusters that can appear both in syllable-initial (onset) position and in syllable-final position (coda), but that differ in their relative probability of occurring in a particular position, were tested as possible boundary cues. Some indications were found that people are sensitive to sequential probabilities. However, no effects were found that strongly support the hypothesis that people do indeed use these distributional properties of the lexicon in the segmentation of spoken language.

INTRODUCTION

In written language, the spaces between words unambiguously indicate word boundaries, but in speech such a clear demarcation does not exist. Nevertheless, listeners hear speech as a series of words. How is speech segmented? Possible boundary cues are acoustic-phonetic cues, metrical cues and phonotactics. In the literature we find evidence for all these boundary cues, and although these cues are not always available, listeners seem to use them when they do occur. Acoustic-phonetic cues in English are, for instance, lengthening of onset syllables and segments (Gow & Gordon, 1995, Lehiste, 1972) and aspiration of word-initial stops (Lehiste, 1960). A large amount of crosslinguistic research has been dedicated to the study of metrical cues for segmentation; listeners are able to use the dominant rhythmic structure of their native language to help locate word boundaries (Cutler & Norris, 1988, for English; Vroomen & de Gelder, 1995, for Dutch; Mehler, Dommergues, Frauenfelder & Segui, 1981 and Cutler, Mehler, Norris & Segui, 1986 for French; Otake, Hatano, Cutler & Mehler, 1993 for Japanese).

Another possible cue for word boundaries is provided by the phonotactic permissibility of sound sequences; if people perceive a combination of sounds that is not allowed within a syllable or word in their language, they may assume a boundary between those sounds. McQueen (1998) used a word-spotting task (Cutler & Norris, 1988) to investigate the role in segmentation of syllable boundaries which are mandatory on phonotactic grounds. He found that detection of the Dutch word *rok* (skirt) is much harder in *fie-drok*, where the target is misaligned with the syllable boundary of the bisyllabic nonword (voiced stops like [d] must be syllable-initial in Dutch), than in *fiem-rok*, where the target is aligned with the syllable boundary forced by the phonotactic rule that [mr] is an illegal consonant cluster in Dutch. Similar results were obtained by Dumay and collaborators (Dumay, Banel, Frauenfelder & Content, 1998; Dumay, Content & Frauenfelder, 1999) in a word-spotting task in French.

If listeners can judge the legality of a perceived combination of sounds in their native language and hypothesize possible word boundaries in continuous speech according to those phonotactics, perhaps they are also able to use less extreme sequential probabilities of their language. In Dutch there are lots of words beginning with the cluster [sp], but only very few words starting [ks]. Conversely there are lots of words that end with the cluster [ts], while there are very few words ending [sk]. Do people use these distributional properties of the lexicon in the segmentation of continuous speech? This is the general question addressed in the present investigation. Hearing a combination of sounds that often occurs immediately before a syllable or word boundary may facilitate recognition of a following word but may interfere with recognition of a preceding word; hearing a combination of sounds that occurs often immediately after a syllable or word boundary may facilitate recognition of a preceding word but may interfere with recognition of a following word. A word-spotting experiment was designed to test this hypothesis. Will it be easier to detect the Dutch word *bloem* (flower) in *bloem-spuum* than in *bloem-ksuum*? Will it be harder to spot *vriend* (friend) in *duusk-vriend* than in *duuts-vriend*?

EXPERIMENT 2.1

Method

Materials

In this experiment, six clusters ([sk], [sp], [st], [ks], [ps] and [ts]) from the complete set of 21 types that occur both as onsets and as codas in Dutch, were selected. The six clusters are optimal with respect to two important properties: first, they are minimally different with respect to acoustic features; and second, they cover the entire range of onset/coda ratios that are found in the analysis of the Celex computerised database of Dutch (Burnage, 1990). All occurrences of consonants and consonant clusters as onsets and codas of syllables and of words were counted, both by type and by token. From this list the 21 types were selected that can occur both as onsets and as codas in Dutch. For these 21 types relative frequency of occurrence was calculated as the ratio of the number of onset occurrences by the number of coda occurrences. Figure 2-1 shows the relative frequencies of occurrence by type for syllables and for words. The selected six consonant clusters range from onset dominant (e.g. the cluster [sk] is 7.5 times more likely as an onset than as a coda) to coda dominant (e.g. the cluster [ks] is 6.8 times more likely as a coda than as an onset).

Each cluster was placed in the onset of a nonsense syllable, after an initial target word (*bloem*, flower, in, for example, *bloem-ksuum* or *bloem-skuum*) and in the coda of a nonsense syllable, before a final target word (*vriend*, friend, in, for example, *duuks-vriend* or *duusk-vriend*). The target words were selected in a way that there was, as far as possible, a phonotactically mandatory syllable boundary between the target word and the cluster.

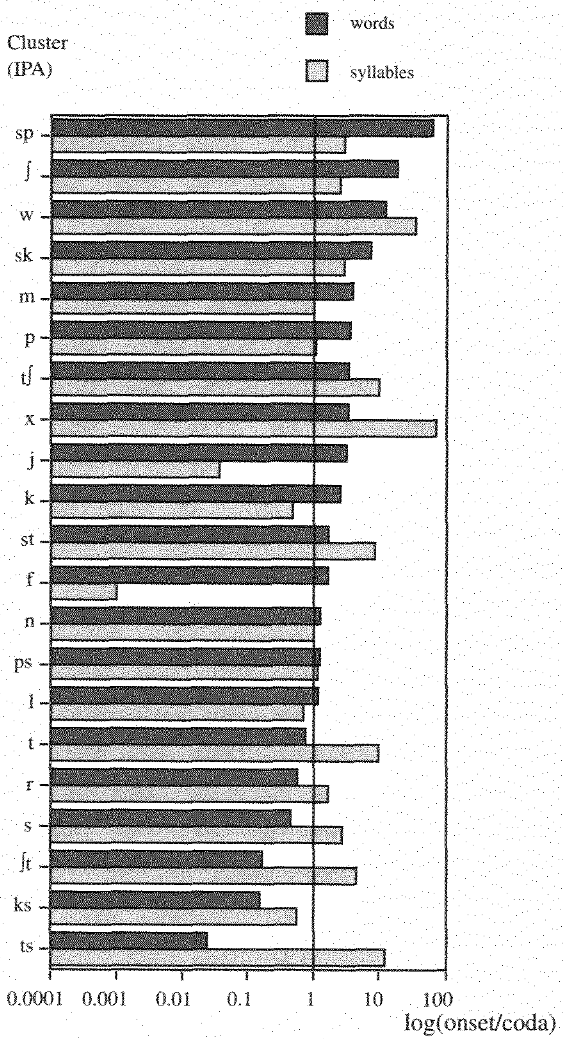

Figure 2-1 *Relative frequency of occurrence of consonants and consonant clusters as onsets and codas of syllables and words. Onset/coda ratios are given on a logarithmic scale by type. An onset/coda ratio larger than one means that a consonant (cluster) is onset dominant, a ratio smaller than one means that a cluster is coda dominant.*

To obtain a large enough set of materials some target contexts were used that did not form a phonotactically illegal sequence with the cluster or the adjacent phoneme of the cluster in Dutch; in those cases the cluster (or the phoneme of the cluster that was adjacent to the target) and the target context never occur in a single rime (e.g. [øːnst] in *deun-staam*p) or in a single onset (e.g. [nst]) in Dutch.

All subjects listened to a total of 96 target-bearing bisyllables (48 initial and 48 final, 8 for each cluster) and 192 bisyllabic fillers with no embedded words. All of the target words were paired with all of the six clusters (see also Table 2-1). The six target-cluster pairs were then counterbalanced across six different versions of the experiment, in such a way that subjects in each group heard a different cluster in the context of each particular target word.

Initial Targets	Final Targets	Relative Probability
bloem-**spuum**	duu**sp**-vriend	Onset Dominant
bloem-**skuum**	duu**sk**-vriend	↓
bloem-**stuum**	duu**st**-vriend	
bloem-**psuum**	duu**ps**-vriend	
bloem-**tsuum**	duu**ts**-vriend	
bloem-**ksuum**	duu**ks**-vriend	Coda Dominant

Table 2-1 *Example materials*

Recording

The items were recorded onto DAT tape in a sound attenuated booth. All items were spoken three times in random order by a female native speaker of Dutch. After the recording the stimuli were redigitized onto a computer. The materials were measured and spliced into separate speechfiles, using the Xwaves speech editor. For each item the most natural utterance was selected. The target items were inspected to make sure that the target-bearing bisyllables were all syllabified correctly between target word and cluster. The individual speechfiles, one for each nonsense bisyllable, were then transferred to the hard disk of a personal computer for use in the experiment.

Subjects

Seventy-two subjects from the Max Planck subject pool participated. Each of the six different versions of the experiment was heard by twelve subjects. All of the participants were native listeners of Dutch. None of them reported any hearing disorder. They were paid for participation.

Procedure

The presentation of the stimuli, the timing of the manual responses and the collection of the data were performed by the NESU experimental system, which was developed at the Max Planck Institute for Psycholinguistics. Subjects were tested two at a time in individual sound attenuated booths. They were asked to press a response button with their preferred hand whenever they heard a real word embedded in one of the bisyllables, and to say which word they heard. Their answers were monitored on-line for false alarms, that is, trials where subjects spotted a word other than the intended target word.

Results and discussion

All manual responses that were accompanied by an incorrect spoken response were set to zero and treated as missing responses, in the same manner as for the trials where subjects failed to respond. Subjects did not respond at all in 20.8% of the trials with target-bearing bisyllables, and they detected words that were not the intended targets in 2.9% of these trials. Responses of duration less than 200 ms or greater than 1800 ms were also treated as missing data. A total of 15 items was excluded from further statistical analysis, because subjects failed to respond correctly to those targets in more than 50% of the cases.

Reaction times (RTs) were measured from target word offset. The RTs were then submitted to Analyses of Variance (ANOVAs) with both subject (*F1*) and items (*F2*) as the repeated measure. The analysis of the RTs showed a highly significant effect of target position, $F1(1, 66) = 404.98$, $p < .001$, $MSE = 20{,}998$; $F2(1, 81) = 215.05$, $p < .001$, $MSE = 37{,}149$. Subjects responded faster to final targets (526 ms) than to initial targets (724 ms). This pattern of results was also found in the error analysis. Subjects made more errors on initial (23.8%) than on final targets (17.6%). This overall effect was confirmed by the ANOVAs on the error data ($F1(1, 66) = 26.77$, $p < .001$, $MSE = .031$; $F2(1, 81) = 7.02$, $p < .01$, $MSE = .099$). Posthoc pairwise comparisons using t tests for correlated means by subjects for the six separate Cluster Conditions revealed reliable differences in the RTs to initial and final targets for all six clusters (RT by subjects, [ks]: $t1(71) = 13.31$, $p < .001$; [ps]: $t1(71) = 13.21$, $p < .001$; [ts]: $t1(71) = 9.77$, $p < .001$; [sk]: $t1(71) = 12.38$, $p < .001$; [sp]: $t1(71) = 14.08$, $p < .001$; [st]: $t1(71) = 14.77$, $p > .001$). The pairwise comparisons between the two target positions per Cluster Condition in the error rates showed that the advantage for final targets was only reliable for three of the clusters (error rate by subjects, [ks]: $t1(71) = 3.69$, $p < .01$; [ps]: $p > .1$; [ts]: $t1(71) = 3.73$, $p < .001$; [sk]: $t1(71) = 4.38$, $p < .001$; [sp]: $p > .1$; [st]: $p > .1$). Because different target words were used in initial and final position no pairwise comparisons using t tests were performed by items.The advantage for final targets is in line with earlier findings (cf. McQueen, 1998) that initial targets are harder to detect in a mixed design with both initial and final targets.

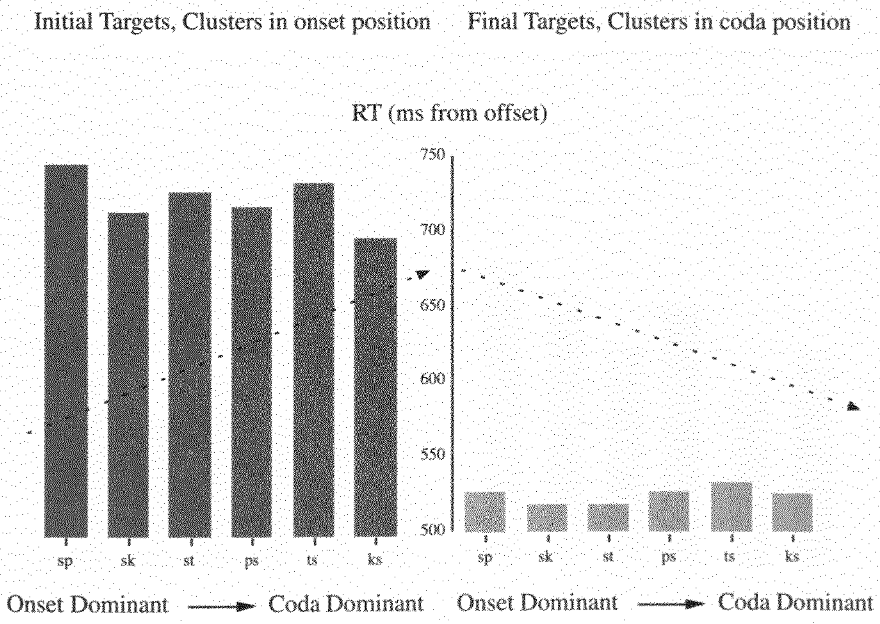

Figure 2-2 *Mean Reaction Times (RTs, in msec, from word offset) from subject analysis for the six clusters in initial and final position in Experiment 2.1. The predicted trends are depicted by the dashed lines.*

No clear effects were found that support the hypothesis that word recognition in this task was influenced by the relative frequency of either following or preceding clusters. Because the predicted effect of a given cluster would be opposite in the two different target positions, the interaction of a cluster with target position is the source of variation that is relevant here ($F1(5{,}330) = 1.64$, $p > .1$, $MSE = 5{,}780$; $F2(5{,}405) = 0.65$, $p > .1$, $MSE = 5{,}620$; see Figure 2-2 for the mean reaction times in the subjects analysis). Similar effects were found measuring the response latencies from word onsets (as reported in van der Lugt, 1997).

Although no overall effects of Cluster Condition in the reaction time data were significant, there were trends in the error data (see Figure 2-3 for the mean error rates in the subjects analysis). For example, the error analysis by subjects showed that subjects made most errors on initial targets that are followed by the cluster [ks] (.30 relative to an overall mean for initial targets of .24), and subjects made most errors on final targets that are preceded by the cluster [sp] (.26 relative to an overall mean for final targets of .18).

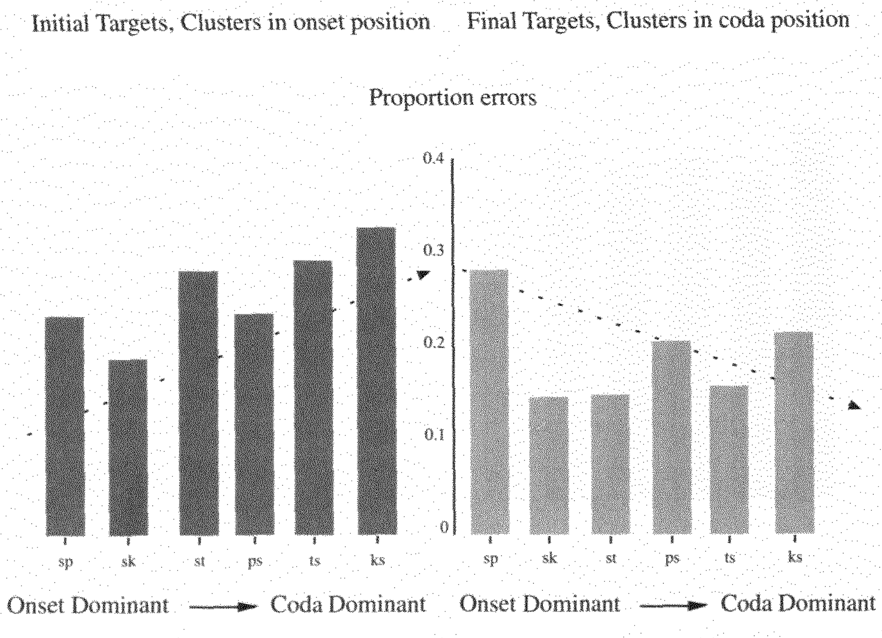

Figure 2-3 *Mean error rates from subject analysis for the six clusters in initial and final position. The predicted trends are depicted by the dashed lines.*

These effects are in line with the predictions: [ks] is the most coda-dominant cluster, and [sp] is the most onset-dominant cluster. The error analyses show a significant overall effect of the relative probability of occurrence of the consonant clusters in a certain position on the proportion of errors ($F1(5,330) = 6.62, p < .01, MSE = .024$; $F2(5,405) = 2.77, p < 0.05, MSE = .032$). Posthoc pairwise comparisons between all 15 pairs of the six Cluster Conditions using *t* tests for correlated means by subjects and items for the two positions separately revealed that only one pair was reliably different using a Bonferroni adjusted alpha level ($p < .003$). Significantly more errors were made to targets in initial position that were followed by the cluster [ks] in syllable onset position than to the clusters [sp] and [st] in onset position ([ks] versus [sp]: $t1(71) = 3.74, p < .003$; [ks] versus [st]: $t1(71) = 3.18, p < .003$). These differences, however, were only significant by subjects (both *t2*s, $p >.01$). Thus, although there was some support in the error data for the hypothesis that sequential probabilities of clusters would influence word-spotting performance, there was certainly no unambiguous support for this assumption.

In this experiment one might expect that the conditions with a [s] directly preceding or following the target are more difficult than the conditions with a stop ([k],[p],[t]) adjacent to the target. There would be two plausible explanations for such an effect. An acoustic-phonetic explanation would be that recognition of a target is facilitated when it is followed or preceded by a stop, because stops are acoustically more salient than fricatives. A second explanation would be a phonological account in the same direction: targets following or preceding an [s] may be harder to recognise due to the role of the [s] in complex consonant clusters; the [s] has a high probability of occurrence both as an onset and as a coda. The data were reanalysed in terms of the phoneme next to the target, stop versus [s]. It was found that subjects' detection latencies did not differ significantly in these two conditions ($F1(1,66) = 1.77, p > 0.1$; $F2(1,247) = 3.82, p>0.5$).

GENERAL DISCUSSION

There are two possible interpretations for these results. A strong interpretation of this experiment would be that people do not use sequential probabilities in the segmentation of speech. The other interpretation would be that the trends found in the error data are indicative of a weak sensitivity to sequential probabilities, and that the experiment lacked sufficient power for these effects to reach statistical significance. A series of follow-up experiments (Chapter 3) were therefore designed, in order to control for some factors that were not controlled for in this first experiment.

The predicted effects of the relative probability of a cluster as an onset or as a coda might have been obscured by effects of the absolute frequencies of occurrence in both positions. Moreover, the complexity of the syllable boundaries and, related to that, possible ambiguities in syllable boundaries because of the role of the segment [s] in complex consonant clusters in Dutch (the [s] has a high probability of occurrence in clusters, both as an onset and as a coda) may have also interfered with the predicted effects of relative probability. In the follow-up experiments, the possible use of the absolute probabilities of CV and VC sequences as boundary cues was studied in word-spotting tasks similar to that used here. However, instead of a range of relative probabilities, two strictly bimodal sets of low versus high probability CV onsets (e.g., *dou* versus *di* in *boom-douf* versus *boom-dif*, with the target *boom*, tree) and low versus high probability VC codas (e.g., *uul* versus *eel* in *buul-veer* versus *beel-veer*, with the target *veer*, feather) were tested as possible boundary cues in the context of the same target word.

The results from the present experiment suggest that people are sensitive to the distributional properties of the lexicon that were investigated, that is, the relative probability of occurrence of complex consonant clusters at the beginnings and ends of segments. These effects were, however, very small and inconsistent. The present results therefore do not provide us with strong evidence that this kind of information is used by listeners in the segmentation of spoken language.

THE USE OF SEQUENTIAL PROBABILITIES IN THE SEGMENTATION OF SPEECH

CHAPTER 3

(manuscript submitted for publication in *Perception & Psychophysics*)

A. H. van der Lugt

ABSTRACT

The present investigation addresses the possible utility of sequential probabilities in the segmentation of spoken language. In a series of five word-spotting and two control lexical decision experiments, high versus low probability Consonant-Vowel (Experiments 3.1, 3.2, 3.5 & 3.7) and Vowel-Consonant (Experiments 3.1, 3.3, 3.4 & 3.6) strings were presented either in the nonsense contexts of target words (Experiments 3.1-3.3) or within the target words themselves (Experiments 3.4-3.7). The results suggest that listeners indeed use sequential probabilities as a cue for segmentation.The probability of a sound sequence influenced segmentation more when the sequence occurred within the target words (Experiments 3.4-3.7 versus Experiments 3.1-3.3). Furthermore, the effects were stronger when the sequences occurred in onset position (Experiments 3.1, 3.2, 3.5 & 3.7 versus Experiments 3.1, 3.3, 3.4 & 3.6).

INTRODUCTION

“To wreck a nice beach" or "To recognise speech?” Ambiguities like this, where the latter utterance can be mistaken for the former, show that the boundaries between words are not reliably marked in speech (Klatt, 1976; Lehiste, 1972; Nakatani & Dukes, 1977; Quené, 1992). However, in order to understand spoken language, the listener has to segment the continuous speech stream into discrete units that convey the meaning of an utterance, that is, into words. People use several strategies to deal with this problem of segmentation.

Although spoken language does not contain the clear demarcation of word boundaries that is provided by spaces in written language, there are several sources of information that can be used, when available, to parse the speech stream into words. Long silences, acoustic-phonetic cues (Gow & Gordon, 1995; Lehiste, 1972; Nakatani & Dukes, 1977; Quené 1989, 1992, 1993), phonological restrictions (e.g. vowel harmony in Finnish: Suomi, McQueen & Cutler, 1997; Vroomen, Tuomainen & de Gelder, 1998), phonotactics (McQueen, 1998) and metrical cues (Cutler & Norris, 1988 for English; Vroomen & de Gelder, 1995 for Dutch; Cutler, Mehler, Norris & Segui, 1986 for French; Otake, Hatano, Cutler & Mehler, 1993 for Japanese) all provide information on the likely location of word boundaries. More importantly, there is evidence that human listeners use all of these boundary cues if they are available in the speech signal and, furthermore, that they seem to use the cues that are most effective in particular situations. If, for example, a stream of speech contains both phonotactic and metrical cues to the location of a word boundary, the adult listener will rely more strongly on the phonotactic information (McQueen, 1998). Mattys, Jusczyk, Luce, & Morgan (1999), however, found the reverse to be true for 9-month olds. They demonstrated that infants rely more strongly on the prosodic information that is available in the speech signal.

But this is not the whole story. Segmentation can arise as a consequence of word recognition: If you know the word, you know the word boundaries. This maxim is the core assumption in two influential models of spoken word recognition: TRACE (Elman & McClelland, 1988; McClelland 1979, 1987, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981) and SHORTLIST (Norris, 1994; Norris, McQueen, Cutler & Butterfield, 1997). Both models avoid explicit segmentation procedures by allowing the initial activation of multiple word candidates and the subsequent selection of the appropriate candidate from that set. Segmentation occurs as a product of this process. The models settle on an optimal parse of the input, that is, a sequence of words with boundaries between them, and achieve segmentation and recognition at the same time. In TRACE and SHORTLIST the selection of the appropriate candidate is the result of competition among activated word candidates in memory. This lexical solution to the problem of segmentation can be integrated with the use of explicit boundary cues in the signal. Through the operation of the "Possible Word Constraint", candidates that are misaligned with likely boundaries are disfavored (Norris et al., 1997).

But how can you find the boundaries, if you don't know the words? How does an infant, learning a language, segment fluent speech in the absence of a lexicon of candidates that can compete for recognition? Studies in language acquisition show that, like adults, when infants listen to their native language they are sensitive both to prosodic regularities (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Svenkerud, & Jusczyk, 1993; Turk, Jusczyk, & Gerken, 1995) and phonotactic legality constraints (Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Friederici & Wessels, 1993), suggesting that these cues can be exploited by the infant in the segmentation of speech.

Phonotactics constitute the extreme, categorical case of judging the familiarity of a perceived combination of sounds: is this combination of sounds legal in this language? However, there is evidence that infants are also sensitive to more gradient transitional probabilities. Jusczyk, Luce and Charles-Luce (1994) found that by the age of nine months, infants prefer patterns of sounds that are relatively common in their native language over combinations of sounds that are relatively uncommon, but not illegal. A recent study by Mattys et al. (1999) examined 9-month olds' sensitivity to within- and between-word transitional probabilities in speech segmentation and how this interacted with prosodic information. They demonstrated that infants prefer sequences that are very common at word juncture when the prosodic and pause information is consistent with a two-unit structure, and that they prefer sequences that have a high probability of occurring inside words when this information is consistent with a single unit. Several computational demonstrations of word segmentation in acquisition (Brent & Cartwright, 1996; Cairns, Shillcock, Chater, & Levy, 1997; Christiansen, Allen, & Seidenberg, 1998) have also confirmed the utility of sequential probabilities as cues to the location of word boundaries.

Sequential probabilities also affect adult speech perception. Saffran and collaborators have shown that both infants (Aslin, Saffran & Newport, 1998; Saffran, Aslin & Newport, 1996) and adults (Saffran, Newport & Aslin, 1996) can utilise transitional probabilities in the segmentation of an artificial language. Vitevitch, Luce, Charles-Luce and Kemmerer (1997) found that subjects' goodness ratings of spoken nonsense words vary as a function of the sequential likelihood of the segments that they were composed of, and they replicated this finding in a speeded auditory repetition task, showing that it was easier for the participants to repeat nonwords composed of frequent segments. Using the same repetition task, Vitevitch and Luce (1998) demonstrated that nonwords composed of phonological segments with high-probability sequential patterns are responded to more quickly and accurately than nonwords composed of phonological segments with low-probability patterns. Evidence that listeners use sequential probabilities in making phonetic decisions was found in phoneme-monitoring experiments (Pitt & Samuel, 1995; McQueen & Pitt, 1996) and in phonetic categorization experiments (Pitt & McQueen, 1998). In summary, these studies show that adult listeners are sensitive to transitional probabilities, but so far no experiment has explicitly asked if and how these distributional cues are exploited by adult listeners in the segmentation of natural speech. The present experiments were designed to address exactly that question.

Experiments 3.1-3.5 all used the word-spotting task. Experiments 3.1-3.3 were designed to investigate the use of Consonant-Vowel probabilities and Vowel-Consonant probabilities in the context of target words. In Experiment 3.1, the target words could appear both in initial and final position. The results show that listeners found it harder to spot words that were preceded by a high-probability Vowel-Consonant sequence. Experiments 3.2 and 3.3 examined whether the observed differences in Experiment 3.1 were due to the intermixing of initial and final targets. It was shown that when all words were embedded at the same position within a single experiment, no reliable differences could be observed between low and high probability sequences, neither in the preceding nor in the following context. Experiments 3.4 and 3.5 investigated the use of similar sequential probabilities of sequences within the target words. The outcomes of these last two experiments, together with the results from two control lexical decision experiments (Experiments 3.6 and 3.7), suggest that sequential probabilities are most valuable for segmentation when they cue the onsets of words.

EXPERIMENT 3.1

In Experiment 3.1 listeners were asked to listen to a list of bisyllabic nonsense items and to try to spot embedded words either at the beginnings or ends of the stimuli. In Dutch there are many words that begin with [dɪ] but only a few words that start with [dɑu]. It was hypothesized that it would be easier for a listener to detect *boom* (tree), for example, in *boom-dif*, where the Dutch target word *boom* is followed by the high probability Consonant-Vowel (CV) onset [dɪ], than in *boom-douf*, where the same target is followed by the low probability CV onset [dɑu]. Similarly, many words in Dutch end with the sequence [e:l], but only a few words end with the sequence [y:l]. It was predicted that it would be easier to spot the Dutch word *veer* (feather) when preceded by a high probability Vowel-Consonant (VC) offset, for example, *veer* in *beel-veer*, than when preceded by a low probability VC, for instance *veer* in *buul-veer*.

Method

General Method Experiments 3.1-3.5

In Experiments 3.1-3.5 the same task and procedure were used and similar general criteria for the selection of materials were adopted. The experiments differed in design, specific stimulus materials and the identity of the participants. In this section all the identical features will be described.

Sequential probability criteria. Initially, all absolute frequencies of occurrence for all onset/nucleus combinations (1048 pairs) and all nucleus/coda combinations[1] (1390 pairs) were computed using the phonological transcriptions in the CELEX computerised database of Dutch (Burnage, 1990).[2] These sets were then restricted to all non-complex pairs, thus yielding all the diphone combinations with a vowel in the Dutch language.

---

[1] *Onset, nucleus and coda are used to describe the structure of syllables in phonology (e.g. Booij, 1995). Each syllable consists of an obligatory vocalic part (the nucleus), preceded by zero or more consonants (the onset) and followed by zero or more consonants (the coda).*

Probability	Sequence			
	CV		VC	
	mean	SD	mean	SD
Low				
Syllable (Type)	40.7	39.5	26.1	29.1
Syllable (Token)	607.6	774.5	309.4	457.3
Word (Type)	14	13.57	7.5	7.6
Word (Token)	188.6	257.0	172.5	208.9
High				
Syllable (Type)	7,573.0	8,982.6	7,949.9	9,998.7
Syllable (Token)	460,687.9	428,788.4	555,797.4	499,503.6
Word (Type)	2,720.2	3,209.5	2,566.6	3,366.2
Word (Token)	314,432.9	268,098.5	404,040.5	361,242.0
Note: corpus size is 42 million tokens				

Table 3-1 *Frequency of occurrence of the experimental sequences*

For every diphone combination (383 CV pairs and 300 VC pairs) four different absolute frequency counts were collected: The number of occurrences at the beginning (CV) or end (VC) of a syllable in a word type and of a syllable in a word token (weighted with the word frequencies in 42 million words), and the number of occurrences at the edges (beginnings or endings) of word types and tokens. The two sets were then further restricted by including only those pairs that have a frequency larger than zero in the word type count, thus filtering out all pairs that never occur as the onset or offset of a word. The

---

2. At present, there is no corpus of spoken Dutch available for the direct computation of frequencies of occurrence of diphones in speech. Instead, diphone frequencies from the CELEX corpus of written language were computed using the phonological transcription of each word and the written word frequency. Correlation analyses between the non-zero diphone probabilities taken from the English CELEX lexicon and those taken from the Marsec corpus of spoken English (Roach, Knowles, Varadi, & Arnfield, 1993) have shown that the CELEX diphone frequencies are reasonable estimates of spoken frequencies ($\underline{r}(1237) = .74$, $\underline{p} < .001$).

remaining 334 CV pairs and 192 VC pairs were then consecutively ranked by all of the four frequency counts. Low probability diphones fell in the intersection[3] of the four lowest quarters of the ranked lists (63 CV and 40 VC candidates), high probability diphones fell in the intersection of the four highest quarters (39 CV and 25 VC candidates). See Table 3-1 for the mean frequencies and standard deviations; a full listing is given in Appendix 3-A.

Materials. Bisyllabic nonsense words were constructed, where either the initial syllable or the final syllable was a real word in Dutch that contained no other embedded words and where the other syllable (final or initial) was a nonword. Diphone pairs from the four probability sets were used either within the target words (Experiments 3.4 and 3.5) or in the nonsense contexts (Experiments 3.1-3.3). The target words were selected in such a way that there was a phonotactically mandatory syllable boundary between the target word and the nonsense syllable as often as possible (7 exceptions in a total of 80 target-bearing bisyllables in Experiments 3.1-3.3; 9 exceptions in a total of 72 target-bearing bisyllables in Experiments 3.4 and 3.5). To obtain large enough sets of materials, some target contexts were used that do not form a phonotactically illegal sequence across the syllable boundary in Dutch; in those cases the sequence has a very low probability of occurrence within a syllable (not in more than 100 tokens in 42 million). Examples of materials for all experiments are given in Table 3-2. Sets of fillers with no embedded words were created for all experiments.

Recording. All of the stimuli were recorded in a sound attenuated booth. All items were spoken three times in random order by a female native speaker of

---

[3] *The strongest facilitatory effect in segmentation would be expected for strings that are frequent at word edges and infrequent word-internally (as was pointed out by John Kingston). However, typically both by type and token, syllable and word position-specific frequencies of CV and VC strings are highly correlated. Furthermore, instances where word and syllable frequencies diverge tend to have a specific linguistic status. High syllable frequencies combined with low word frequencies are typical for CV strings that are common final syllables, like plural inflections in Dutch. Conversely, low syllable frequencies with high word frequencies are typical for CV strings that are common first syllables, for instance the past participle prefix [xə]. Asymmetries between type and token counts are mostly caused by CV and VC segments that form function words in Dutch, like [də]. The use of the intersection of the four sets was motivated by this observation. Since the main aim was to investigate the role of sequential probabilities, confounds with higher level linguistic information, for instance, the potential role of function words and morphemes in segmentation, were avoided.*

Dutch. After the recording, all items were redigitised onto a computer at a sampling frequency of 16 kHz. All the materials were measured and cut into separate speech files, using the Xwaves/ESPS speech signal processing package. For each item the most natural utterance was selected. The target items were inspected to make sure that all target-bearing bisyllables were syllabified correctly between target word and nonsense syllable. The speech files were then transferred to the hard disk of a personal computer for use in the experiment.

Sequential probability	Target position	
	Initial	Final
in **context** sequence	Experiments 3.1+3.2	Experiments 3.1+3.3
Low	*boom* (tree) in *boom-**douf***	*veer* (feather) in ***buul**-veer*
High	*boom* (tree) in *boom-**dif***	*veer* (feather) in ***beel**-veer*
within word sequence	Experiment 3.4	Experiment 3.5
Low	*heup* (hip) in ***heup**-juik*	*geur* (smell) in *pien-**geur***
High	*kap* (cap) in ***kap**-juif*	*galg* (gallows) in *hien-**galg***

Table 3-2 *Design materials and examples, Experiments 3.1-3.5*

Subjects. The participants in all 7 experiments were members of the subject pool of the Max Planck Institute for Psycholinguistics. None of the them reported any hearing problems and they were all native listeners of Dutch. An equal number of subjects participated in the two different versions of each experiment.

Procedure. The stimuli were presented binaurally via Sennheiser headphones. The presentation of the stimuli, the timing of the manual responses and the collection of the data was performed by the NESU experimental control software. Subjects were tested two at a time in individual sound-attenuated booths. The participants were instructed to listen to a series of bisyllabic nonsense words. They were asked to press a response button with their preferred hand whenever they heard a real word embedded in one of the bisyllables, and to say aloud which word they had heard. Their answers were monitored on-line for false alarms (trials where they spotted a word other than an intended target word).

Method experiment 3.1

Materials. Eight CV onset+nucleus pairs, four from the high and four from the low sequential probability set were paired with five different C(C) codas to serve as following target contexts and, similarly, four VC nucleus+coda pairs from both sets were paired with five different (C)C onsets to serve as preceding contexts. The consonants of the four critical diphones of the low and high probability sets were matched as far as possible on the identity of the consonant. In those cases where matching was impossible, the consonants shared manner of articulation and only differed in voicing or place of articulation. This matching was necessary to make the acoustic salience of the syllable boundary as constant as possible across the two different probability conditions. The non-critical VC(C) and (C)CV diphone pairs were all selected from within the second and third quarters of the ranked (C)CV and VC(C) sequential probabilities, these portions of the contexts were therefore neither high nor low in frequency. Two sets of 80 target-bearing nonsense bisyllables were constructed, one set consisting of 40 initial CVC target words followed by 40 nonsensical CVC(C) contexts, 20 in each condition, and 40 final CVC target words preceded by 40 nonsensical (C)CVC contexts, again 20 in each condition. The two different target-context pairs for every target were counterbalanced across two different versions of the experiment, so that both versions contained a different context for each particular target word. A target word could appear only once in a single version of the experiment, combined either with a high probability context or with a low probability context. Each version therefore contained a total of 80 target bearing bisyllables (40 initial and 40 final, 20 for each probability condition) and a set of 160 fillers (128 CVC-CVC, 16 CCVC-CVC and 16 CVC-CVCC) with no embedded words. All 80 target words had a frequency in the range 100-1000 in 42 million tokens of the CELEX computerised database of Dutch; the items are listed in Appendix 3-B.

Subjects and procedure. Forty subjects participated in Experiment 3.1.

Results and discussion

Data analysis, all experiments

False alarms were set to zero and treated as missing responses, as were the trials in which subjects failed to respond entirely. Responses of duration less than 200 ms or greater than 1800 ms were also treated as missing data. Reaction times were measured from (target) word offset. The reaction times (RTs) and the error rates were submitted to Analyses of Variance (ANOVAs) with both subjects (*F1*) and items (*F2*) as the repeated measure. For the RT analyses means were computed per condition, by subjects and by items separately, disregarding the errors.

Results Experiment 3.1

As shown in Table 3-3, subjects found it harder to spot words that were preceded by a probable word-final context (e.g. *veer* in *beel-veer*) than to detect words that were preceded by an improbable word-final context (e.g. *veer* in *buul-veer*). On the other hand, words that were followed by a very common word-initial context (e.g. *boom* in *boom-dif*) were not detected much easier than words followed by a very uncommon word-initial context (e.g. *boom* in *boom-douf*).

Sequential probability	Target position	
	Initial	Final
Low		
RT (ms)	593	393
Error rate (%)	12.4	12.6
Dutch example	*boom* (tree) in *boom*-***douf***	*veer* (feather) in ***buul***-*veer*
High		
RT (ms)	567	475
Error rate (%)	11.3	21.9
Dutch example	*boom* (tree) in *boom*-***dif***	*veer* (feather) in ***beel***-*veer*

Table 3-3 *Mean reaction times and error rates, Experiment 3.1*

This pattern of results was observed in the RT analyses. There was a significant effect of boundary probability, *F1*(1, 38) = 6.03, $p < .05$, *MSE* = 5,028; *F2*(1, 78) = 4.58, $p < .05$, *MSE* = 7,108, and a highly significant interaction of target position and boundary probability, *F1*(1, 38) = 29.10, $p < .001$, *MSE* = 4,046; *F2*(1, 78) = 21.09, $p < .001$, *MSE* = 7,108. This pattern was also found in the error analyses: a significant effect of boundary probability, *F1*(1, 38) = 10.31, $p < .01$, *MSE*= .01; *F2*(1, 78) = 5.41, $p < .05$, *MSE* = .01, and a highly significant interaction of target position and boundary probability, *F1*(1, 38) = 6.03, $p < .05$, *MSE* = .01; *F2*(1, 78) = 8.81, $p < .01$, *MSE* = .01. Overall, the responses to targets in the high probability condition were slower than in the low probability conditions and subjects made more errors on high than on low probability items. Since there was in fact an effect in the opposite direction for initial targets, these main effects of boundary probability are misleading. The interaction was inspected more closely by performing *t* tests. Pairwise comparisons showed that the only reliable differences in RTs and errors were between the high and low probability condition for final targets, RT by subjects *t1*(39) = 4.33, $p < .001$; RT by items *t2*(39) = 4.26, $p <. 001$, errors by subjects *t1*(39) = 5.15, $p < .001$; errors by items *t2*(39) = 2.94, $p <. 01$. For the initial targets the differences between the high and low probability in RT were only marginally significant, RT by subjects *t1*(39) = -2.03, $p < .05$; RT by items *t2*(39) = -2.00, $.01 < p <. 05$. The errors were not different (both *p*'s>.1).

On average, subjects responded 146 ms faster to final targets than to initial targets. This effect of target position in the RTs was highly significant, *F1*(1, 38) = 92.85, $p < .001$, *MSE* = 9,154; *F2*(1, 78) = 30.31, $p < .001$, *MSE* = 29,263. This is in line with earlier findings that initial targets are harder to spot than final targets in a mixed design with both initial and final targets (McQueen, Norris, & Cutler, 1994). This pattern was repeated only partly in the error analysis: it was significant by subjects (*F1*(1, 38) = 17.70, $p < .001$, *MSE* = .01) but not by items (*F2*(1, 78) = 2.47, $p > .1$, *MSE* = .05).

The observed effects do not strongly support the initial hypothesis that it should have been easier for listeners to detect words that were followed by a very likely beginning and contradict the prediction that it would be easier to spot words that were preceded by a very likely ending. However, there is a potential problem with the particular task of Experiment 3.1. The listeners were asked to detect target words that could appear in either the initial or the final syllable of the stimuli. A possible explanation of the observed effect could be that the high-probability word-final contexts before the final targets were suggestive of the presence of an initial target, thus slowing down the detection of the actual finally-embedded target word. The highly probable VC sequences activate many lexical candidates in the initial position. These candidates all compete with each other for recognition, but are more suggestive of an initial target than the fewer lexical candidates activated by the lower probability VC sequences. Detection of the final target is thus delayed in the high probability condition. This explanation is in line with findings by Luce et al. on the role of similarity neighbourhood frequencies in spoken word recognition (e.g. Cluff & Luce, 1990; Goldinger, Luce & Pisoni, 1989; Luce & Pisoni, 1998; Luce, Pisoni & Goldinger, 1990). Furthermore, Vitevitch and Luce (1998, 1999) have shown that sequential probability manipulations can have opposite effects on word recognition performance depending on the nature of the processing environment and the task demands: inhibitory lexical competition effects and facilitatory sequential probability effects. The finding that high-probability contexts made detection of final targets harder could thus be a lexical competition effect, one that was emphasized by task-level competition between initial and final targets.

The absence of such a competition effect for initial targets can be explained by the sequential nature of spoken word recognition: the later incoming information of the following context is too late to interfere with the detection of the initial target. Effects of following context on word-spotting performance have been shown to be weak (McQueen, 1998), or even absent (Suomi et al., 1997).

To explore these issues, two experiments were carried out with subsets of the materials of Experiment 3.1. In these experiments subjects monitored for target words in only one position. In a series of word spotting experiments McQueen et al. (1994) have shown that blocked target location can attenuate lexical effects. It was therefore predicted that if the listener could attend to a single location, the beginning or the end of a nonsense bisyllable, response-level competition with lexical candidates in the other position would decrease. Subjects were asked to listen only for initially embedded words in Experiment 3.2, and only for finally embedded words in Experiment 3.3.

EXPERIMENT 3.2

Method

Materials. All 80 initial target-bearing nonsense bisyllables from Experiment 3.1 were mixed with 80 appropriate fillers. Again, the two different target-context pairs for every target were counterbalanced across two different versions of the experiment, so that both versions contained a different context for each particular target word. Each version therefore contained a total of 40 initial target bearing bisyllables (20 for each probability condition) and a set of 80 fillers (64 CVC-CVC, 16 CVC-CVCC) with no embedded words.

Subjects and procedure. Forty subjects participated in Experiment 3.2.

Sequential probability	Target position	
	Initial (Experiment 3.2)	Final (Experiment 3.3)
Low		
RT (ms)	506	363
Error rate (%)	6.6	8.9
Dutch example	*boom* (tree) in *boom-**douf***	*veer* (feather) in ***buul**-veer*
High		
RT (ms)	481	373
Error rate (%)	6.1	8.3
Dutch example	*boom* (tree) in *boom-**dif***	*veer* (feather) in ***beel**-veer*

Table 3-4 *Mean reaction times and error rates, Experiments 3.2 and 3.3*

Results and discussion

Subjects were on average 25 ms faster in spotting targets that were followed by a high probability boundary CV (see Table 3-4). This difference, however, was not completely reliable. The analyses of RT revealed a significant effect of boundary probability in the subjects analysis, $F1(1, 38) = 7.29$, $p < .05$, $MSE =$

1,728, but this effect did not reach significance in the items analysis, $F2(1, 78) = 1.88$, $p > .1$, $MSE = 5{,}292$. No reliable differences in listeners' performance were found in the error analyses.

The results of Experiment 3.2 are in line with the results for the initial targets of Experiment 3.1. Listeners responded faster to words that were followed by a very likely word onset sequence, but not reliably so. No differences were found in the error rates. Although these results are suggestive of the predicted facilitatory effect of a common following context, they do not provide strong evidence that position-specific sequential probabilities are used by the listener in this situation.

EXPERIMENT 3.3

Method

Materials. The 80 final target-bearing nonsense bisyllables from Experiment 3.1 were combined with 80 appropriate fillers in two different versions. The design was identical to that in Experiment 3.2.

Subjects and procedure. Forty subjects participated in Experiment 3.3.

Results and discussion

Subjects were on average 10 ms faster in detecting embedded words that were preceded by a low probability boundary VC (see Table 3-4). This difference, however, was not statistically reliable. Both in RT and in error rates no reliable differences ($p > .1$) were found between the two probability conditions.

In Experiment 3.1, in which subjects were asked to listen for words that were embedded either at the beginnings or at the ends of the bisyllabic nonsense items, it was harder for listeners to detect words that were preceded by a probable word-final context than to detect words that were preceded by an improbable word-final context. In the present experiment subjects were monitoring exclusively for final embedded words in a set of materials that did not contain any initial embedded words. Under these circumstances it was equally hard to detect a target in a common preceding context as in an uncommon preceding context. This is in line with the prediction that the attentional focus on a single location (i.e. the end of the stimulus) would diminish the effects of competition with candidates in the other location (i.e., the beginning of the stimulus).

Nevertheless, no actual reversal of the lexical competition effect was observed. This is not surprising. Too a large extent competition is an automatic feature of spoken word recognition (see also McQueen et al., 1994).The high probability sequences will still have activated many lexical candidates causing a weaker competition effect, even when the possibility of embedded words in the initial position was not emphasized by the task or the materials. This lexical competition effect may still have cancelled out any facilitatory segmentation effect of sequential probability. A second series of experiments was therefore planned where sequential likelihoods were manipulated *within* the embedded words. In order to avoid effects that could be caused by the intermixing of initial and final targets, these effects were investigated in two separate word-spotting experiments with initial and final embeddings respectively.

EXPERIMENT 3.4

In Experiment 3.4 listeners were asked to listen for embedded words at the beginnings of bisyllabic nonsense items. Hearing a combination of sounds that often occurs immediately before a syllable or word boundary, e.g. [ap] in Dutch, may facilitate recognition of a target word that ends with that combination of sounds. Conversely, hearing a combination of sounds that seldom occurs as the end of a syllable or word, e.g. [øːp] in Dutch, may interfere with the recognition of a word that ends with that sequence. It was hypothesized that it would be easier to detect *kap* (cap) in *kap-juif*, than to detect *heup* (hip) in *heup-juik*.

Method

Materials. A slightly different procedure was used for the construction of materials from the sets of CV and VC segments. First, a list was generated that consisted of all the monosyllabic words of Dutch that end in one of the VC nucleus+coda pairs from the low sequential probability set. Secondly, a similar list was constructed for the high probability set. The words from these lists were matched on their syllabic structure, the manner of articulation of the final consonant and as far as possible on place of articulation of the final consonant. The two sets of words were also matched as far as possible on the frequencies of occurrence of the words in the CELEX database and on the variance of these frequencies[4]. This resulted in two sets of 15 words, 9 CCVC and 6 CVC, with a mean log frequency of 2.14 (*SD* = .50) for the high probability set and a mean log frequency of 2.16 (*SD* = .49) for the low probability set. Two sets of 15 nonsense syllables, without embedded words, were constructed using 15 CV segments with normal positional segment probabilities (i.e. segments selected from within the second and third quarters of the ranked CV and VC sequential probability lists) that were all paired with two different coda consonants (see also Table 2). These 15 pairs of nonsense syllables were placed in the final context of the target words in such a way that the two target words that were matched on their final consonant were combined with a context that started with the same CV segment. The experiment therefore contained a total of 30 target-bearing bisyllables (15 for each probability condition) and a set of 60 fillers (24 CVC-CVC and 36 CCVC-CVC) with no embedded words. The target-bearing items are listed in Appendix 3-C.

Subjects and procedure. Twenty-one subjects participated in Experiment 3.4.

Results and discussion

Inferential statistics, Experiments 3.4-3.7. Because different items were used in the two probability conditions, additional analyses were performed to investigate possible effects of the specific items that were used. Correlation analyses and Analyses of Covariance (ANCOVAs) are reported for target duration and/or target frequency only if those differences indeed had an effect (as revealed by a significant correlation).

*Results, Experiment 3.4.*The results suggest that it is easier to detect words with a common VC offset. Listeners were faster and made less mistakes when the embedded words ended with a very likely sequence (such as [ap] rather than [øːp]).

---

[4] *At a later stage subjective frequency ratings (on a scale from 1 to 7) were collected from 30 subjects for all the target words that were used in Experiments 4, 5, 6 and 7. These ratings were reasonably correlated with the CELEX frequencies (*r(72)=.60, p < .001*, and, more importantly, the targets were well balanced in each experiment (Experiments 4+6: High Probability mean rating = 4.4,* SD=*1.4, Low probability mean rating=4.5,* SD=*1.3; Experiments 5+7: High Probability mean rating = 4.5,* SD=*1.1, Low probability mean rating=4.3,* SD=*1.1).*

Sequential probability	Target position	
	Initial (Experiment 3.4)	Final (Experiment 3.5)
Low		
RT (ms)	669	492
Error rate (%)	19.7	8.9
Dutch example	*heup* (hip) in ***heup**-juik*	*geur* (smell) in *pien-**geur***
High		
RT (ms)	620	444
Error rate (%)	5.2	7.4
Dutch example	*kap* (cap) in ***kap**-juif*	*galg* (gallows) in *hien-**galg***

Table 3-5 *Mean reaction times and error rates, Experiments 3.4 and 3.5*

Subjects were on average 49 ms faster in spotting targets that ended in a high probability boundary VC (see Table 3-5). The RT analyses revealed a significant effect of boundary probability in the subjects analysis, $F1\ (1, 21) = 5.77$, $p < .05$, $MSE = 4{,}397$, but not in the items analysis, $F2(1, 28) = 2.87$, $p > .1$, $MSE = 13{,}090$. In the analysis of the errors, more reliable differences in listeners' performance were found between the probability conditions. Listeners made on average 14% more errors in spotting targets that ended in a low probability boundary VC. This effect was significant both by subjects, $F1(1, 21) = 48.58$, $p < .01$, $MSE = .001$, and by items, $F2(1, 28) = 7.00$, $p <. 05$, $MSE = .02$.

Although an attempt was made to control for differences in target frequency and variation targets in the High Probability set had, on average, a slightly smaller log frequency than the words in the Low Probability set. A significant negative correlation of target log frequency and RT was found ($r(30) = -.41$, $p < .05$). An ANCOVA with target frequency as covariate showed that the effect of probability was more reliable than in the items ANOVA, but it remained only marginally significant, $F2(1, 27) = 3.67$, $p <. 1$, $MSE = 11{,}005$.

EXPERIMENT 3.5

In Experiment 3.5 listeners were asked to listen exclusively for embedded words at the ends of bisyllabic nonsense items. Hearing a combination of sounds that often occurs immediately after a syllable or word boundary, for instance [xa] in Dutch, may facilitate recognition of a target word that begins with that combination of sounds; hearing a combination of sounds that seldom occurs as the beginning of a syllable or word, e.g. [xøː] in Dutch, may interfere with the recognition of a word that begins with that sequence. It was hypothesized that it would be harder to spot *geur* (smell) in *pien-geur*, than to spot *galg* (gallows) in *hien-galg*.

Method

Materials. The materials for this experiment were constructed in a similar way to those in Experiment 3.4. This resulted in two sets of 21 words, 4 CVCC and 17 CVC, with a mean log frequency of 2.30 (*SD*=.56) for the high probability set and a mean log frequency of 2.28 (*SD*=.54) for the low probability set. Two sets of 21 nonsense syllables, without embedded words, were constructed using 21 VC segments with normal positional segment probabilities that were all paired with two different onset consonants. These 21 pairs of nonsense syllables were placed in the initial context of the target words in such a way that the two target words that were matched on their initial consonant were combined with a context that ended with the same VC segment (see also Table 3-2). The experiment therefore contained a total of 42 target-bearing bisyllables (21 for each probability condition) and a set of 84 fillers (16 CVC-CVCC and 68 CVC-CVC) with no words embedded in them. The target-bearing items are listed in Appendix 3-C.

Subjects. Twenty-one subjects participated in Experiment 3.5.

Results and discussion

Subjects were on average 48 ms faster in spotting targets that started with a high probability boundary CV (see Table 3-5). The RT analyses revealed a highly significant effect of boundary probability by subjects, $F1(1, 21) = 10.58$, $p < .01$, $MSE = 2{,}357$, but not by items, $F2(1, 40) = 1.90$, $p > .1$, $MSE = 12{,}145$. Listeners made just as many errors in spotting targets that began with a common CV as in spotting targets with an uncommon CV onset (*F1* & *F2 ns*).

Targets in the High Probability set were shorter than targets in the Low Probability set (Low Probability 497 ms, High Probability 454 ms) and the targets in the High Probability set had, on average, a slightly higher log frequency than the words in the Low Probability set. Further analyses showed a significant negative correlation of target duration and RT ($r(42) = -.42$, $p < .01$) and a significant negative correlation of target frequency and RT ($r(42) = -.41$, $p < .01$). These findings are consistent with results from previous word-spotting studies. Subjects have been shown to respond faster to longer embedded targets (McQueen, 1998). The present results, however, cannot be explained by durational differences between the two probability conditions. In fact, the target duration effect is working against the sequential probability effect, since the items in the High Probability set were shorter than those in the Low Probability set. It has also been shown that listeners are faster and more accurate in spotting high-frequency than low-frequency words (Freedman, 1992). In the present study the items from the High Probability set were slightly more frequent than the items from the Low Probability set. Thus, part of the processing advantage for High Probability items might be explained by this difference in frequency. Frequency and duration were therefore both entered as covariates in an ANCOVA of the RTs by items. The effect of probability was now significant, $F2(1, 38) = 6.63$, $p < .05$, $MSE = 8{,}207$. This suggests that differences in duration and frequency between the High Probability and the Low probability items were partly neutralising the effect of positional probability. Similar analyses were performed for the errors. Both duration ($r(42) = -.51$, $p < .01$) and frequency ($r(42) = -.49$, $p < .01$) were significantly correlated with the error rates. An ANCOVA showed no significant differences between the two probability conditions.

These results suggest that the faster detection latencies for words that began with a common CV onset cannot be explained by durational differences or frequency differences between the High and the Low probability sets. But since different targets were used in these two probability sets, it was necessary to make a closer examination of the possibility that differences other than the probability of the initial CV segments between the two sets were responsible for this effect. As control experiments, lexical decision experiments were used to investigate the possibility that differences other than the segment probabilities of the target words made it easier to spot words that ended (Experiment 3.4) or started (Experiment 3.5) with common segments. In order to keep the experiments as closely related as possible to the earlier experiments separate control experiments were carried out for the embedded words from Experiment 3.4 and 3.5.

EXPERIMENT 3.6

If the different effects for the Low versus High Probability target words from the word-spotting experiments are due to the relative ease of segmentation that is caused by the manipulated probabilistic cues to the location of the syllable boundary, than these differences should not be found if the target words are presented in isolation. It is assumed that when monosyllabic words and nonwords are presented in isolation, that is when the beginnings and ends of the words are unambiguously cued by silence, no effects of sequential probability will arise, because there is no need for segmentation in this situation. But if the relative ease of spotting targets from the High Probability set was caused by some other difference that was not directly related to the ease of segmentation (for example an acoustic difference between the items in the High and the Low Probability set) the same pattern of results should be found in the lexical decision task.

In Experiment 3.6, all of the initially embedded target words from Experiment 3.4 were cut out from the original bisyllabic nonsense item. Nonsense monosyllables were constructed by cutting out the first syllables from the Experiment 3.4 bisyllabic fillers. All items were presented to the listener in the same order as was used in Experiment 3.4, and they were asked to respond only when they heard a real word by pressing the response button and to say aloud the real word that they heard. This go/no-go lexical decision task is a fairly standard control for word-spotting experiments (McQueen 1996).

Method

Materials. The items were all constructed from the items that were used in Experiment 3.4. Target words were constructed by excising the first syllables from the target-bearing bisyllables; the first syllables from the original filler items were cut out to serve as nonword fillers. The items were excised at zero-crossings on the basis of visual and auditory inspection of the waveform under the constraint that this point was consistent across the two probability conditions. For this purpose the Xwaves speech editor was used. The order of presentation was the same as that in Experiment 3.4.

Subjects and procedure. Twenty-four subjects participated. Subjects that had already participated in Experiments 3.4 or 3.5 were excluded. The procedure was similar to the procedure used in the word spotting experiments. The only difference was that participants were now instructed to respond whenever they heard a real word. As before, subjects were asked to name the word they recognised after the manual response.

Results and discussion

Listeners were on average 23 ms faster in detecting targets that ended in a high probability boundary VC (see Table 3-6), but this difference was not significant (both p's $> .1$). In the analysis of the errors a more reliable difference was found in the listener's performance. Listeners made 10 % more errors in detecting targets that ended in a low probability boundary VC. This effect was highly significant by subjects, $F1(1, 23) = 30.19$, $p < .001$, $MSE < .01$, and marginally significant by items, $F2(1, 28) = 3.65$, $p > .05$, $MSE = .02$.

Additional correlation analyses were performed in order to quantify the relationship between these results and the word spotting results. Lexical decision and word-spotting latencies were highly correlated ($r(30) = .60$, $p < .001$) and also the error rates show a robust correlation ($r(30) = .64$, $p < .001$). As a final control, the RT data and the error data from the lexical decision experiment were entered as covariates in ANCOVAs of the word-spotting responses of Experiment 3.4. These analyses showed that when the effects of the specific target items used were filtered out, no reliable differences were left between the detection of embedded words that end with a likely VC offset and the detection of embedded words that end in an unlikely VC offset (RT: $F2(1, 27) = 2.30$, $p > .1$, $MSE = 8{,}797$; Errors: $F2(1, 27) = 3.04$, $p > .05$, $MSE = .02$).

EXPERIMENT 3.7

In Experiment 3.7, all of the finally embedded target words from Experiment 3.5 were cut out from the original bisyllabic nonsense item. Nonsense monosyllables were constructed by cutting out the second syllables from the Experiment 3.5 fillers.

Method

Materials. The items were all constructed from the items that were used in Experiment 3.5, in the same way as in Experiment 3.6.

Subjects and procedure. Twenty-four subjects participated. Subjects that had already participated in Experiment 3.4, 3.5 or 3.6 were excluded. The procedure was the same as in Experiment 3.6.

Sequential probability	Target position	
	Initial (Experiment 3.6)	Final (Experiment 3.7)
Low		
RT (ms)	506	463
Error rate (%)	18	24.2
Dutch example	*heup* (hip) excised from ***heup**-juik*	*geur* (smell) excised from *pien-**geu**r*
High		
RT (ms)	483	487
Error rate (%)	7.5	26.2
Dutch example	*kap* (cap) excised from ***kap**-juif*	*galg* (gallows) excised from *hien-**ga**lg*

Table 3-6 *Mean reaction times and error rates, Experiments 3.6 and 3.7*

Results and discussion

Listeners were on average 24 ms slower in detecting words that started with a high probability boundary CV (see Table 3-6). This difference was not significant ($F1$ & $F2$, $p > .1$). The subjects never recognised words other than the intended targets, but the error rate was much higher than in the original word-spotting experiment (25.2% versus 8.1%). There is a simple acoustic explanation for this increase in errors. Because the finally embedded target words were excised from their original context, many of the beginnings of the items were not very clear due to residual coarticulatory information from the offset of the first syllable. However, this affected the items in both probability sets equally. In the analysis of the errors also no reliable differences were found (both *p*'s > .1).

Lexical decision and word-spotting latencies were again highly correlated ($r(42) = .63$, $p < .001$), and the error rates from both tasks again showed a robust correlation ($r(42) = .51$, $p <. 001$). The ANCOVA of the word-spotting errors with the lexical decision errors as covariate yielded no reliable differences between the two probability conditions. The ANCOVA of the word-spotting RTs with the lexical decision RTs as covariate confirmed the earlier findings that it was easier for the listener to detect words that begin with a common onset than words that begin with an uncommon onset ($F2(1, 39) = 5.33$, $p <. 05$, $MSE = 6{,}964$). It is therefore unlikely that the observed effect for the common onsets in Experiment 3.5 was due to the use of particular target items in the two probability sets.

GENERAL DISCUSSION

These experiments have systematically examined the influence of sequential probabilities on the segmentation of speech. In Experiment 3.1 and the control Experiments 3.2 and 3.3 it was demonstrated that the listeners' sensitivity to sequential likelihoods can be obscured by higher order effects due to competition between lexical candidates. When targets could occur in both initial and final position (Experiment 3.1) the sequences of segments in the context activate lexical candidates to the extent that they interfered with detection of the intended target word. In line with McQueen et al. (1994; Experiment 2) this competition effect was shown to be in part task-specific. In Experiments 3.2 and 3.3, where target words occurred in a single position, the lexical competition effect was no longer observed. This is also consistent with recent results of studies by Vitevitch and Luce (1998, 1999). They have demonstrated that the processing of speech is a function of both facilitatory effects associated with sequential probabilities and inhibitory effects associated with lexical competitors. In particular, they have shown that the task environment can differentially emphasize lexical and sublexical levels of processing.

In the second series of experiments sequential probability was manipulated within the target words. The results of Experiment 3.4 and 3.6 provide no reliable support for the hypothesis that listeners use the likelihood of a word's *offset* as a boundary cue when this boundary is also marked by a phonotactically mandatory boundary. Experiment 3.5 and the control lexical decision Experiment 3.7, however, demonstrated that the listener can exploit the likelihood of a word's *onset* in combination with a phonotactic boundary in the segmentation of speech. Common sequences at the onset of a word made it easier for the listeners to detect these words in a word spotting task; however, they did not affect performance in the lexical decision task. These findings provide converging evidence that sublexical levels of processing operate in the recognition of spoken stimuli, even for words. In line with previous research (Vitevitch and Luce, 1999) it has been demonstrated that sublexical sequential probability effects manifest themselves when effects of lexical competition are minimised. In Experiment 3.4 and 3.5 of the present study this was established by manipulating sequential probabilities within the target words, and by blocking target position.

The likelihood of a word's onset seems to be more important than the likelihood of a word's offset. Many studies have confirmed the intuitive notion that the beginnings of words are particularly important for the fast recognition of words in continuous speech. Gating studies (Grosjean, 1980; Tyler & Wessels, 1983, 1985) have shown that words can be recognised reliably before their offset, at the point where they become unique from all other words in the language (the "uniqueness point"). Other evidence from phoneme monitoring (Marslen-Wilson, 1984), auditory lexical decision (Taft, 1986), mispronunciation detection (Cole, 1973; Marslen-Wilson & Welsh, 1978) and cross-modal priming (Marslen-Wilson, 1987; Marslen-Wilson & Zwitserlood, 1989; Zwitserlood, 1989) also suggests that word-initial information is crucial for efficient word recognition. In the present study there is a confound between increases in word spotting latency and the position of the target word. The reaction times for initial target words in Experiment 3.4 were much longer than reaction times for final targets in Experiment 3.5. Two issues are relevant with respect to this difference. First, statistical power may have been reduced by the higher variance associated with longer reaction times. Second, if effects of sequential probability are short-lived, one might argue that subjects responded too late for probability effects to be observed. Therefore, no strong conclusions on the positional specificity of the sequential probability effect can be drawn on the basis of these results.

In Experiments 3.4 and 3.5, words with high probability sequences by definition come from high density similarity neighborhoods. Similarly, words with low probability sequences come from a low density neighborhood. Therefore, the findings in Experiments 3.6 and 3.7 could be interpreted as a failure to replicate the results of Luce & Pisoni (1998), whose Neighborhood

Activation Model predicts that words in dense lexical neighborhoods should take longer to recognize than words in sparse neighborhoods. However, as was also shown by Vitevitch and Luce (1999; Experiment 2), lexical effects can be substantially attenuated by the task demands and processing environment. In the present experiments this attenuation of lexical effects was most likely due to the fact that the words were excised from a bisyllabic context.

These findings provide further evidence that listeners have access to information regarding the probability that a particular sequence of phonemes occurs at a certain position within a syllable or word. More importantly, the results also show that listeners exploit this kind of information in the segmentation of words from longer sequences. Current implementations of models of spoken word recognition, for instance the SHORTLIST model (Norris et al., 1997), lack a mechanism to account for these findings. Such a mechanism would incorporate some kind of prelexical analysis of the speech signal that would result in the accumulation and storage of distributional information. A different approach is taken by Vitevitch and Luce (1999). They have proposed an adaptive resonance framework (Grossberg, Boardman, & Cohen, 1997) to account for both the lexical and sublexical level of processing speech. In this approach the two levels of processing are modelled by a single mechanism (rather than a prelexical and a lexical mechanism). Resonant states between chunks in short term memory and items in working memory constitute the speech percept. If sublexical chunks (corresponding to units smaller than words) in short-term memory are the most predictive (for instance, in the case of a nonword input), the model reflects sublexical processing. If lexical chunks are consistent with the item in working memory (for instance, when the input is a word), they will dominate processing since they represent a larger stretch of the input. In this case, the model reflects a lexical level of processing (see also Vitevitch & Luce, 1999).

A final point to note about these results is that the use of sequential probability information as a cue to segmentation has only been investigated in conjunction with one other cue to the location of word boundaries, namely those that are marked by phonotactics. The different findings for adults (McQueen, 1998) and infants (Mattys et al., 1999) on the integration of prosodic and phonotactic boundary markers show that the operation of an explicit boundary cue can vary enormously depending on the situation. Future research should therefore include experiments on the interaction of sequential probability information before and after the boundary, the interaction with other boundary cues, and the use of sequential probability cues in the absence of other explicit boundary cues.

In summary, the results suggest that listeners indeed use sequential probabilities as a cue for segmentation. Two limitations were observed. The probability of a sound sequence influenced segmentation more when the sequence occurred within the target words. Furthermore, the effects were stronger when the sequences occurred in onset position.

COMPETITION AND MISMATCH IN SPOKEN WORD RECOGNITION

CHAPTER 4

(manuscript submitted for publication in *Journal of Experimental Psychology: Human Perception and Performance*)

A. H. van der Lugt

ABSTRACT

How tolerant is spoken word recognition with respect to word initial mismatching information? McQueen, Norris and Cutler (1994) found that it is harder to spot a word like *mess* in [dəmɛs], which is the beginning of *domestic*, than in [nəmɛs] which is not the beginning of an English word. This competition effect was here utilised in two identity priming experiments and two word-spotting experiments with three critical conditions: a word onset condition (e.g., [xo:lɛs] which is the beginning of the Dutch word *cholesterol*); a minimal mismatch condition where the initial phoneme only differed from the word in place of articulation (e.g., [fo:lɛs]), and a maximal mismatch condition where the initial phoneme differed in place, manner and voice (e.g.,[bo:lɛs]). Overall, the results showed that the word recognition system is rather intolerant of word-initial phonetic mismatches. However, the results also suggest that mismatches involving only a single feature are less distinctive than mismatches involving several features.

INTRODUCTION

What happens if a drunk asks you for a *shigarette* instead of a *cigarette*? In other words, how tolerant is the word recognition system with respect to mismatching information? Norris (1982, 1994) argued that models of spoken word recognition should be able to deal with this kind of mispronunciation. However, what is the evidence that human listeners can and do deal with mismatching information in the speech signal?

Subphonetic mismatches

Subphonetic mismatches, as defined by Whalen (1984), are artificial mismatches where inappropriate cues are inserted in the input that are not sufficient to change the phonetic percept. Whalen found that when an [s] or an [ʃ] is combined with vocalic formant transition information that is appropriate to a different fricative, phonetic judgments are slower than when there is no mismatch. Whalen also showed that phonetic decisions on syllable initial stops [p] and [k] are slowed down by following mismatching vowel information. Subsequently, Whalen (1991) investigated the effect of subcategorical mismatches in a lexical decision task and a naming task. He found that lexical decision responses were slowed down for both words (e.g. *soup*) and nonwords (e.g. *shoup*) when the listeners were presented with misleading cues. In the auditory naming task only the slower responses were affected by mismatching information. These results suggest that phonetic mismatches are resolved before lexical access. However, an earlier lexical decision experiment by Streeter and Nigro (1979) did show an interaction between lexical status and appropriateness of the cues. They found that lexical decision responses to words were slowed down by mismatching information, whereas the responses to nonwords were not affected. A possible explanation for these contradictory results (see also Whalen, 1984) might be that a large proportion of the nonwords that were used by Whalen were extremely like real words (e.g. *shoup*), such that the misleading cue, although still subcategorical, was pointing toward a word rather than a nonword. Whalen found that responses to these nonwords were significantly slower than the responses to nonwords that were less like any existing words.

A more recent study by Andruski, Blumstein, and Burton (1994) examined the effects of subphonetic mismatches on lexical activation using a unimodal semantic priming paradigm. They systematically shortened the Voice Onset

Time (VOT) of the initial voiceless consonant (e.g. [k]) of an auditory prime (e.g. *king*) by removing one-third or two-thirds of the full VOT, thus altering the initial consonant towards its voiced counterpart (e.g. [g]) but such that it would still be categorised as voiceless. The three versions of each prime (unaltered, -1/3 VOT, & -2/3 VOT) were presented just before a related auditory target (e.g. *queen*). Subjects were asked to perform a lexical decision task on these targets. At an Inter Stimulus Interval (ISI) between prime and target of 50 ms, Andruski et al. found significant priming in all three conditions relative to the baseline condition where the same target (e.g. *queen*) was preceded by a semantically unrelated prime (e.g. *bell*). More importantly, they found that the lexical decision response was faster for the unaltered and -1/3 VOT primes than for the -2/3 primes. In addition to this graded priming effect as a function of the amount of mismatch they also found that listeners responded more slowly to targets that were preceded by primes with a real-word voiced rhyming counterpart (e.g. [pɛt]-[bɛt]) than to targets preceded by primes with a nonword voiced counterpart (e.g. [kɪŋ]-[gɪŋ]). No reliable differences in size of semantic facilitation between any of the priming conditions was found at an ISI of 250 ms. These results seem to confirm the earlier explanation of the discrepancy between the results of Whalen (1991) and those obtained by Streeter and Nigro (1979). Mismatching information affects the performance on nonwords via the activation of other possible word candidates.

This view is further supported by the findings of Marslen-Wilson and Warren (1994) and McQueen, Norris and Cutler (in press). In lexical decision experiments in Dutch (McQueen et al., in press) and English (Marslen-Wilson & Warren, 1994) reliable subphonetic mismatch effects were found for nonwords that were created by cross-splicing the final consonant of a nonword with the first portion of a word (e.g. *smob* created from *smog* and *smob*), but no such effect was found for nonwords that were created from other nonwords (e.g. *smob* created from *smod* and *smob*). The mismatching information, in this case the inappropriate cues in the formant transitions of the vowel preceding the cross-spliced consonants, slowed down lexical decision responses only when the inappropriate cues were pointing toward a word.

In summary, the effects of subphonetic mismatches indicate that relatively small acoustic variations at a subcategorical level can influence the performance of human listeners at the lexical level. It has also been shown that misleading subphonetic cues have a graded effect depending on the amount of mismatch. Furthermore, it has been demonstrated that competing lexical candidates are an important factor in tasks that involve access to the mental lexicon.

Ambiguous segments

Ambiguous segments are a special case in research on the tolerance of the spoken word recognition system, since they present the listener with information that is deviant, but that is neither matching nor mismatching. Marslen-Wilson and collaborators (Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996) used perceptually ambiguous onsets to investigate the sensitivity of the system to deviant information. They manipulated the VOT of word initial stops to obtain ambiguous primes for a cross-modal semantic priming experiment in English. Four different auditory targets, for example, the word *task*, the voiced rhyming nonword *dask* or the ambiguous *t/dask* or the semantically unrelated control *crisp*, were presented auditorily just before presenting a visual target, for example JOB. They found that participants were faster in making a lexical decision response to JOB when they had just heard *task* in comparison to the unrelated control condition. They also found that lexical decision to JOB was equally facilitated by hearing the ambiguous prime *t/dask*, but no facilitation was found for the nonword prime *dask*. In a second stimulus group no facilitation effect was found for primes that were ambiguous between two words. For instance, hearing *plank* facilitated lexical decision to WOOD, but hearing the word rhyme *blank* and the ambiguous rhyme *p/blank* did not influence the response latency. As in the case of subphonetic mismatches, the absence of a facilitatory effect for the ambiguous *p/blank* suggests that competing word candidates play an important role in the processing of mismatching information. A similar point was made by Marslen-Wilson, Moss and van Halen (1996) who also argued that "the presence of priming in the no-competitor condition is indeed a competitor effect" (p. 1388).

Connine, Blasko and Wang (1994) also used a cross-modal priming paradigm to investigate the effects of similar perceptual ambiguities between two words. However, they found that lexical decision to a visual target (e.g. LITTLE) was facilitated by hearing a prime that was ambiguous between two words (e.g. *b/pig*), only one of which was semantically related to the target (e.g. *big*). This contrast might be explained as an effect of the time-course of processing. The response times in all of the experiments by Connine et al. were much slower than the reaction times in the Word Competitor condition of the experiment by Marslen-Wilson and colleagues (mean reaction time 664 ms versus 522 ms). If word rhyme priming effects emerge slowly, processing could have been facilitated at a later point in time while remaining unaffected at shorter response latencies.

To summarise, perceptually ambiguous input can in some cases be tolerated by the word recognition system, as revealed by significant priming of semantic associates. In addition, these studies have confirmed the special role of competing word candidates.

Phonetic mismatches

Most of the previous research on the tolerance of human word recognition has been concerned with mismatches at a phonetic level. Phonetic mismatches are instances where the speech signal deviates from all the words in the mental lexicon by at least one entire phoneme. Many different paradigms (see Bölte, 1997, for a review) have been used to study human performance in the case of mismatching phonetic information: Cross-Modal Semantic priming (Marslen-Wilson & Zwitserlood, 1989; Connine, Blasko & Titone, 1993), Gating (Bölte, 1997), Lexical Decision (Taft & Hambly, 1986), Mispronunciation Detection (Cole, 1973; Marslen-Wilson & Welsh, 1978), Phoneme Monitoring (Connine, Titone, Deelman & Blasko, 1997), Shadowing (Marslen-Wilson & Welsh, 1978), Unimodal Semantic Priming (Marslen-Wilson, Moss, & van Halen, 1996) and Word Reconstruction (van Ooijen, 1996). Typical manipulations of the stimuli include the competitor environment, the perceptual distance of the mismatch and the position of the mismatch in a word. The following overview of the empirical literature on phonetic mismatches is limited to previous work that has addressed word initial deviations by a single phoneme.

Marslen-Wilson and Zwitserlood (1989) tested triplets of an original word, a rhyming word and a rhyming nonword, for example (in English) *battle*, *cattle* and *yattle*, in a cross-modal priming experiment using Dutch materials. Each of these stimuli was presented auditorily immediately followed by the visual presentation of a letter string, for example WAR, to which the participants were asked to make a lexical decision. It was shown that only original words (e.g. *battle*) facilitated the response to a related visual target relative to the response to the same target preceded by an unrelated non-overlapping control (e.g. *packet*). However, the word and nonword rhyme primes were constructed in such a way that the initial consonants were as distinct as possible from the initial consonants of the original word. Furthermore, the competitor environment of the stimuli was not varied systematically (see also Marslen-Wilson, 1993).

In another series of experiments, Marslen-Wilson and collaborators (Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996) continued their investigation of mismatch effects using an intramodal auditory-auditory semantic priming paradigm focusing on perceptual distance and competitor environment. They covaried the phonological distance between a word, for example *tomato*, and a nonword prime, for example *pomato*, with the presence or absence of rhyming lexical competitors. The results show that even small mismatches, that is, an input that is just mismatching by one phonological feature (e.g. *pomato*) is not treated by the system as an instance of the original word. Auditory lexical decision responses to target words that were preceded by the nonword rhyming primes were less facilitated than responses to targets that were preceded by the original words. However, in contrast to the previous results, significant priming effects were obtained for the rhyming nonwords relative to the control condition. Marslen-Wilson et al. explained this difference as a time-course effect. They argued that the slower response times in the unimodal priming experiment allowed priming effects to emerge. In addition, it was shown that the competitor environment influenced performance. Nonword primes with phonologically close rhyming word competitors other than the base word were less effective (26 ms priming effect) than nonword primes without any lexical competitors (57 ms priming effect). Furthermore, no differential effects were found for the amount of mismatch, as defined by the number of mismatching phonological features.

However, there is evidence for graded effects of the amount of phonetic mismatch. Connine, Blasko, and Titone (1993) found that hearing related words and derived minimal nonwords primed lexical decision to a visual target, but no facilitation was found for a derived maximal nonword. For instance, hearing *service* and *zervice* facilitated lexical decision to TENNIS but hearing *gervice* did not affect the response. This finding was confirmed in a more recent study that used a phoneme monitoring task to investigate word-initial mismatch effects (Connine, Titone, Deelman, & Blasko, 1997). It was demonstrated that detection of a final consonant becomes easier as a function of how word-like the stimulus is. For example, detecting [p] in *ketchup* is easier than detecting [p] in *getchup*, detecting [p] in *getchup* is easier than detecting [p] in *retchup*, detecting [p] in *retchup* is easier than detecting [p] in *gidjup*.

The absence versus presence of graded effects of amount of mismatch might be due to the different criteria that were used for the perceptual distance between base words and derived nonwords. An alternative explanation could be given in terms of task differences. Phoneme monitoring might just provide a better, more sensitive measure of mismatch effects than cross-modal semantic priming. Research by Bölte (1997) has shown that graded effects of mismatch can come and go depending on the task that is used. In a series of six unimodal experiments with identical materials but with five different tasks graded effects of the amount of mismatch were found with gating (Grosjean, 1980) and word correction (Bölte, 1997; see also, van Ooijen, 1996). However, no such similarity effects were obtained with auditory lexical decision (Marslen-Wilson, 1980), pseudoword detection (Bölte, 1997) and mispronunciation detection (Cole, 1973).

Competition is a central mechanism of spoken word recognition. A lot of indirect support for competition is provided by studies that show evidence for multiple activation of candidate words (Marslen-Wilson 1987, 1990; Shillcock, 1990; Zwitserlood, 1989; Gow & Gordon, 1995). Phonological priming studies have shown inhibitory effects which probably reflect competition (Goldinger, Luce & Pisoni, 1989; Goldinger, Luce, Pisoni & Marcario, 1992; Slowiaczek & Hamburger, 1992; Radeau, Morais & Dewier, 1989). The most direct evidence for competition comes from studies that used identity priming (Vroomen & de Gelder, 1995) and word spotting (McQueen, Norris & Cutler, 1994; Norris, McQueen & Cutler, 1995). The previous work on subphonetic, ambiguous and phonetic mismatches has also indicated that competition between lexical candidates plays a special role in dealing with mismatching information, but so far no experiment has explicitly investigated the effects of mismatching information on both lexical access and competition. This was the aim of the present investigation. The first two experiments used a cross-modal form priming paradigm (Zwitserlood, 1996), in the last two experiments a word spotting paradigm was used (Cutler & Norris, 1988).

In cross-modal form priming experiments, participants are presented with a visual target, to which a response (for example, lexical decision) is required, preceded by a spoken word or longer utterance that is structurally (as opposed to semantically) related to the visual target in the critical conditions. When visual target and auditory stimulus are related, the auditory stimulus is intended to facilitate or inhibit the processing of the visual target. This effect is measured with respect to a control condition in which the auditory and the visual stimulus are not related. A major advantage of this paradigm is the possibility to present the visual target at various times relative to the presentation of the visual target. This allows one to trace the lexical activation of words in time (cf. Vroomen & de Gelder, 1995).

In the following sections two form priming experiments are reported, in which the participants were asked to perform a different task on a significant proportion of the trials. This 'catch' task was designed to make sure that the participants were also attending to the beginnings of the auditory stimuli. The main aim in these identity priming experiments was to investigate whether differential priming effects, modulated by the amount of mismatch, could be found at different stages in the competition process.

EXPERIMENT 4.1

McQueen, Norris & Cutler (1994) found for English that it is harder to spot a word like *mess* in [dəmɛs] which is the beginning of *domestic*, than in [nəmɛs], which is not the beginning of an English word. This competition effect was utilized in the present identity priming experiment, in Dutch.

In Experiment 4.1 participants were asked to listen to a list of bisyllabic nonsense items. Immediately after hearing the auditory stimulus two things could happen. In the first case a letter string, for example LES, was presented on a computer screen immediately after *choles* had been presented. In this case the participants were asked to decide whether the letter string was an existing Dutch word. In the second case, a warning signal was presented on the screen followed by the visual presentation of a letter string, for example ZUIF after *zuifkaaf*. In this case the participants were asked to decide whether the letter string was an adequate description of the beginning of the nonsense word that they had just heard. This 'catch' task was included to make sure that the participants were also attending to the beginnings of the auditory stimuli.

One would predict, in line with the results obtained by Vroomen and de Gelder (1995), that visual lexical decisions on target words that are preceded by auditory stimuli that have those words embedded in them will be less facilitated when the auditory stimuli also form the onsets of longer words in Dutch. Competition with the longer candidate should lower the activation of the embedded word that is caused by the auditory stimulus, and thereby reduce the priming effect on the visual lexical decision on the embedded word. For instance, it should be harder to make a lexical decision to LES after hearing [xoːlɛs], which is the beginning of *cholesterol*, than after hearing [boːlɛs]. However, what happens in the case of [foːlɛs], where the initial consonant [f] is much closer to the original [x] than the initial consonant [b]? Will it pattern with the word onset condition [xo:lɛs], with [bo:lɛs], or will it be an intermediate case? The results from previous studies suggest that [bo:lɛs] will mismatch with *cholesterol* to block activation of that word, so there will be no competition, and making a lexical decision to LES should be relatively easy. However, previous results do not make clear what is likely to happen with [fo:lɛs].

Method

Materials

Thirty-two polysyllabic words with three or more syllables were selected from the Dutch CELEX lexicon (Burnage, 1990). These words all satisfied the following constraints: neither the first syllable nor the first two syllables together formed an existing Dutch word, but the second syllable was a real word in Dutch. For instance, the first and the first two syllables of the Dutch word *cholesterol* (cholesterol), *cho* and *choles*, are not existing Dutch words, but the second syllable forms the Dutch word *les* (lesson). For every bisyllabic Word Onset (e.g. *choles*) two mismatching versions were created: a Minimal Mismatch condition and a Maximal Mismatch condition. In the minimal change condition the initial consonant of the word onset was changed into a consonant that differed only in place of articulation (e.g. *foles*) in the maximal change condition the initial consonant was substituted by a consonant that differed in place, voicing and manner of articulation (e.g. *boles*). A set of 64 fillers was also made. Half of the fillers, 32, were constructed to be similar to the critical items: 32 polysyllabic words with three or more syllables were selected from the Dutch CELEX lexicon, but now the first two syllables contained no embedded words. Again, two mismatching versions were created for every bisyllabic word onset. This yielded 32 matched filler triplets. For example, the triplet *sugges-fugges-bugges* was constructed from the Dutch word *suggestie* (suggestion). These matched fillers were included to avoid possible strategic responses to word onsets. A complete list of the critical items and the matched fillers is given in Appendix 4-A. The other 32 fillers did not contain embedded words and did not form the onsets of longer words in Dutch. These materials were used as the auditory primes in the present experiment.

The 32 experimental triplets were used to construct quadruplets of prime-target pairs (see also Table 4-1). The embedded words of the triplets, for example *les* from *choles-foles-boles*, were used as visual targets in the three experimental conditions, Word Onset, Minimal Mismatch and Maximal Mismatch and in a control condition. In the control condition the response to the same visual target was measured when it was preceded by a different, unrelated, item from the stimulus set. Half of the control auditory primes were Word Onsets and half of them were Maximal Mismatches.

Filler prime-target pairs were constructed in a similar way. Twelve matched filler triplets were paired with twelve visual nonword targets, six of those were phonologically related to the second syllable of the auditory prime, the other six were unrelated. Phonologically related visual targets shared all phonemes but one with the second syllable of the auditory prime. They were included to avoid participants strategically responding to an overlap between the second syllable of the auditory stimulus and the visual target (see also Vroomen & de Gelder, 1995; Zwitserlood, 1996). Eight matched filler triplets were combined with existing Dutch target words that were phonologically unrelated. The other twelve matched filler triplets served as auditory stimuli on twelve catch trials. Similarly, the free fillers were paired with 20 nonwords, 10 phonologically related and 10 unrelated, and 8 words, the 4 remaining free fillers served as auditory stimuli for the catch trials. A total of sixteen auditory stimuli (12 matched fillers and 4 free fillers) were used for the catch trials. For half of these 16 auditory primes letter strings were created that formed an adequate orthographic description for the first syllable of the auditory prime, for the other 8 auditory primes, letter strings were generated that did not form a possible written version of the first syllable of the prime.

The four different visual target-prime pairs for the experimental items and the matched fillers were counterbalanced across four different versions of the experiment, so that each visual target appeared only once. Each member of an experimental quadruplet or matched filler quadruplet and all free fillers were presented in the same pseudo-random order in all four versions. Each version therefore contained a total of 32 critical pairs (8 in each condition), 20 matched filler pairs (5 in each condition) 28 unrelated filler pairs and a total of 16 catch pairs (12 matched fillers and 4 free fillers). The visual targets were phonologically unrelated to the auditory prime in half of the priming trials, 8 in the control condition of the critical pairs, 14 in the matched fillers and 18 in the free fillers. In the other 40 priming pairs the visual target was either embedded in the auditory prime, i.e. 24 in three experimental conditions, or largely overlapping, 16 in two filler conditions (8 each). The sixteen catch trials were equally spread across the list.

Recording

All of the auditory stimuli were recorded in a sound attenuated booth. All of the bisyllabic items were spoken three times in pseudo-random order by a male native speaker of Dutch. The experimental items and the matched fillers were recorded in triplets, e.g. *choles-foles-boles*, in order to minimize the acoustic differences between the three versions, other than the change of the initial consonant. After the recording all items were digitized at a sampling frequency of 16 kHz. All the materials were measured and cut into separate speech files, using the Xwaves/ESPS speech signal processing package. For each triplet the most constant utterance was selected. The target items were inspected to make sure that all target bearing bisyllables were syllabified correctly between target word and nonsense syllable.

Spoken prime	Visual Target	Condition
choles	LES	Word Onset
foles		Minimal Mismatch
boles		Maximal Mismatch
redak		Control

Table 4-1 *Example materials Experiments 4.1-4.2*

Subjects

Forty-eight subjects participated in Experiment 4.1 in four groups of 12 subjects for each version. Participants were recruited from the Max Planck subject pool. They were paid for participation. No participants reported any vision or hearing disorders.

Procedure

Subjects were tested in groups of four or less. The presentation of the stimuli, the timing of the manual responses and the collection of the data was performed by the NESU experimental system. Participants were each seated in front of a computer screen, in four different compartments of the experimental room that were separated by sound attenuating screens. The auditory stimuli were presented binaurally via Sennheiser headphones. The visual stimuli were presented centered on the computer screen in white against a black background. These letter strings were presented unmasked in 36 point Arial font for 1000 ms. The computer screens were also used to indicate the beginning and the end of the experiment and the pause between the practice block and the experimental block. The participants were instructed to listen to a sequence of nonsense bisyllables. At the offset of each auditory stimulus two things could happen on the computer screen. If a letter string was presented at the offset of the prime, they were asked to perform a lexical decision task on the visual stimuli. They were asked to press a green response button marked "JA" (yes) with their preferred hand as quickly as possible, whenever they saw an existing Dutch word on the screen. If the letter string that was presented on the computer screen was not a real Dutch word, they were asked to press a red button marked "NEE" (no) with their other hand. On the catch-trials a warning signal consisting of three asterisks, "***", was presented visually for one second starting at the offset of the prime. This warning signal was immediately followed by the visual presentation for one second of a letter string. In this case, the participants were asked to press the green response button marked "JA" (yes) with their preferred hand as quickly as possible, if the letter string on the screen formed an adequate description of the beginning of the auditory prime. If the letter string that was presented on the computer screen was not a good description, they were asked to press the red button marked "NEE" (no) with their other hand. The 96 trials of the main experiment were preceded by a block of 9 practice trials.

Results and discussion

First, the performance on the catch trials was analysed separately to investigate if the participants were listening to the beginning of the auditory primes. All of the participants responded correctly to more than 80% of the catch trials, suggesting that they were not ignoring the first part of the auditory primes.

Lexical decision latencies were measured from the onset of the visual targets on the screen. Analyses of variance (ANOVAs) were performed on these decision latencies and on the proportions of errors with both subjects (*F1*) and items (*F2*) as the repeated measure (see Figure 4-1 for the mean responses from the subjects analysis). Non-responses and responses longer than 1200 ms were treated as missing data. The missing latencies were substituted by the mean reaction time per subject per condition for the subjects analysis and by the mean reaction time per item per condition for the items analysis. Listeners were on average 35 ms faster in making lexical decisions to strings that were preceded by a related auditory prime, in comparison to an unrelated control. The differences between the three related conditions, Word Onset, Minimal Mismatch and Maximal Mismatch, were very small.

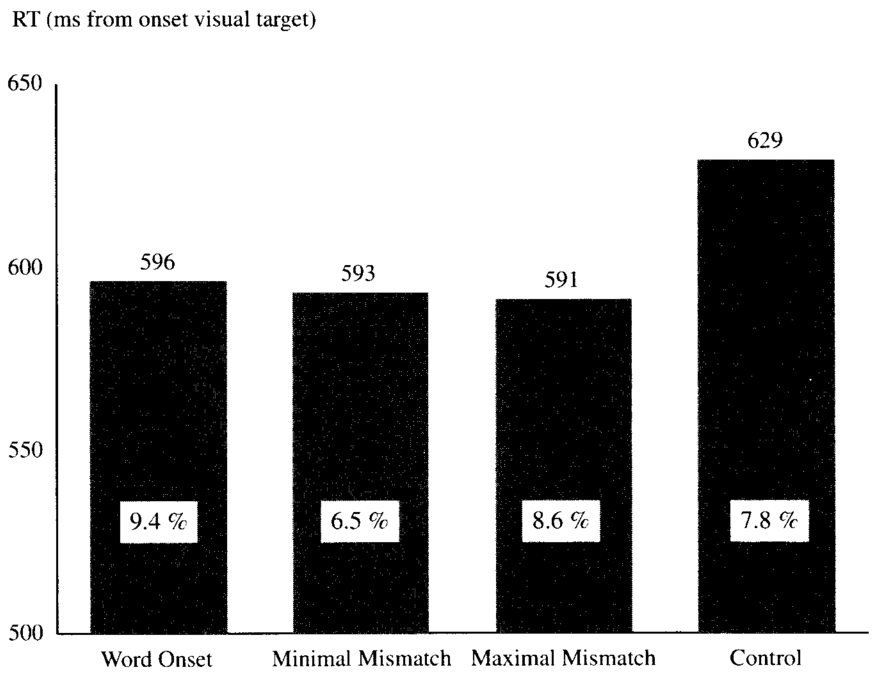

Figure 4-1 *Average lexical decision latencies and error rates (inside the columns) for Experiment 4.1. Results are given for each condition: Word Onset (choles(terol)-LES), Minimal Mismatch (foles-LES), Maximal Mismatch (boles-LES), and Control (redak-LES). Visual targets were presented at the* offset *of the embedded word.*

The ANOVAs of the reaction times suggested that there was an overall effect of priming in the related conditions, $F1(3,132) = 6.28$, $p < .01$, $MSE = 2{,}457$; $F2(3,93) = 3.77$, $p < .05$, $MSE = 3{,}403$. Post-hoc Tukey HSD tests ($p < .05$) and pairwise comparisons confirmed that the Word Onset condition, the Minimal Mismatch and the Maximal Mismatch were not significantly different from each other ($p > .1$), but that all three related conditions were significantly different from the control condition by subjects and by items (Word Onset: $t1(47) = -2.48$, $p < .05$; $t2\ (31) = -2.25$, $p < .05$, Minimal Mismatch: $t1(47) = -3.29$, $p < .01$; $t2(31) = -2.77$, $p < .01$ and Maximal Mismatch: $t1(47) = -3.37$, p

$< .01$; $t2(31) = -3.58$, $p < .01$). In the error data no reliable differences were found between the unrelated control and the three related conditions (*F1* and *F2*, $p > .1$).

Difference scores were computed by subjects and by items to further assess the facilitation by the three different related auditory primes. A difference score by subjects was defined as the mean reaction time of a specific participant to the target words preceded by related primes minus the mean reaction time to the target words preceded by unrelated primes (the baseline control condition) for all three conditions separately. Similarly, the difference score by items was defined as the mean reaction time to a specific target word preceded by a related prime minus the mean decision latency to the same target word preceded by an unrelated prime for all three related conditions separately. The difference scores were again submitted to ANOVAs and post-hoc pairwise comparisons using *t*-tests, both by subjects and by items. These analyses again showed no reliable differences between the three related prime conditions ($p > .1$). The difference scores per priming condition were then tested against zero using *t*-tests[1]. These analyses confirmed that the visual lexical decision to words was facilitated by preceding related auditory primes in all three conditions (Word Onset: $t1(47) = -2.48$, $p < .05$; $t2\ (31) = -2.25$, $p < .05$, Minimal Mismatch: $t1(47) = -3.29$, $p < .01$; $t2(31) = -2.77$, $p < .01$ and Maximal Mismatch: $t1(47) = -3.37$, $p < .01$; $t2(31) = -3.58$, $p < .01$).

In summary, identical priming effects were found for the Word Onset, the Minimal Mismatch and the Maximal Mismatch condition when the visual target is presented at the offset of the embedded word. No differences were found between the three priming conditions. Lexical decision on a visual target is equally facilitated by related auditory primes that form the beginning of a longer word as by related auditory primes that don't. The participants' performance on the secondary catch task suggests that this absence of a differential effect of competition can not be attributed to some kind of attentional strategy, where subjects are ignoring the first part of the auditory stimuli. To investigate whether differential competition effects arise at an earlier stage of processing a second priming experiment was conducted.

---

[1] *These analyses are identical to the posthoc pairwise comparisons for correlated means between the three critical conditions and the control.*

EXPERIMENT 4.2

In Experiment 4.2 the visual stimuli, both for the experimental trials and the 'catch' trials were presented at an earlier point relative to the auditory prime: at the onset of the embedded word.

Method

Materials

The materials were the same that were used in Experiment 4.1. For all of the auditory stimuli the onset of the second syllable was determined by visual and auditory inspection of the waveform. For this purpose the Xwaves speech editor was used. The order of presentation was the same as in Experiment 4.1.

Subjects

Forty-eight subjects from the Max Planck subject pool participated in Experiment 4.2. Twelve subjects were assigned to each of the four versions. Subjects that had already participated in the previous experiment were excluded.

Procedure

The procedure was similar to the procedure used in Experiment 4.1. The only difference was that the visual stimuli, both the letter strings on the identity priming trials and the warning signal on the catch trials, were now presented at the onset of the second syllable.

Results and discussion

Analyses of the performance on the secondary task showed that all subjects correctly responded in more than 80% of the catch trials. This suggests that they were indeed attending to the auditory stimuli.

Lexical decision latencies were measured from the onset of the visual targets on the computer screen. Responses later than 1200 ms and cases where subjects did not respond at all were treated as missing data. Missing data points were again replaced by the relevant mean (per subject or per item) per condition.

The differences between all four conditions, Control, Word Onset, Minimal

Mismatch and Maximal Mismatch, were very small (see Figure 4-2 for the mean results from the subjects analysis). The ANOVAs of the RTs suggested that there is no facilitatory effect of priming ($F1$ and $F2$, $p > .1$). Post-hoc Tukey HSD tests and pairwise comparisons confirmed that the Word Onset condition, the Minimal Mismatch and the Maximal Mismatch were not significantly different from the Control baseline and each other ($p > .1$). Also in the error data no reliable differences were found between the unrelated control and the three related conditions ($F1$ and $F2$, $p > .1$).

As before, difference scores were computed by subjects and by items. The difference scores were again submitted to ANOVAs and post-hoc pairwise comparisons using t-tests, both by subjects and by items. These analyses again showed no reliable differences between the three related prime conditions ($p > .1$). The difference scores per priming condition were then tested against zero using t-tests. These analyses confirmed that the visual lexical decision to words was not facilitated by preceding related auditory primes in any of the three conditions ($p > .1$).

In summary, in Experiment 4.2 no priming effects were found when the visual target was presented at the onset of the embedded word. In Experiment 4.1 equivalent priming was found in all three conditions when the visual target was presented at the offset of the embedded word. In two form priming experiments, Vroomen and De Gelder (1995) demonstrated that lexical decision to the Dutch word MELK (milk) was slower after the spoken prime *melkaam* than after auditory presentation of *melkeum*. This difference was interpreted as evidence for competition between lexical candidates. In Dutch, there are many words that start with *kaa(m)* but only few words that start with *keu(m)*. If the end of the speech input is consistent with many words, lexical decision to an overlapping competitor (e.g. *melk*) is less facilitated than when the final part is consistent with only a few words. Although Vroomen and De Gelder found differential priming effects as a function of the number of overlapping competitors, the present results demonstrate no differential effects of competition.

A possible explanation of the absence of lexical inhibition could be given in terms of the time-course of the competition process. If visual processing starts at the onset of the embedded word in the auditory stimulus (Experiment 4.2), the embedded word (e.g. *les*) might not be activated in any of the three critical conditions and, therefore might not prime visual lexical decision. If visual processing starts at the offset of the embedded word (Experiment 4.1), the embedded word should be activated and prime visual lexical decision in all three conditions, however, the embedding word (e.g. *cholesterol*) could be activated to a similar degree in all three conditions and not give rise to a differential effect of competition.

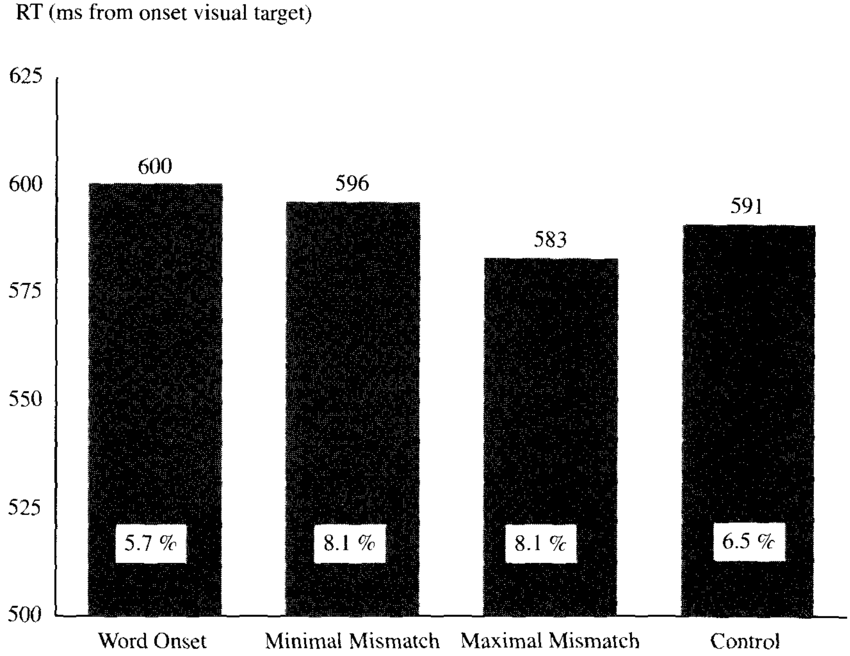

Figure 4-2 *Average lexical decision latencies and error rates (inside the columns) for Experiment 4.2. Results are given for each condition: Word Onset (choles(terol)-LES), Minimal Mismatch (foles-LES), Maximal Mismatch (boles-LES), and Control (redak-LES). Visual targets were presented at the* onset *of the embedded word.*

This possibility was investigated by running simulations with SHORTLIST (Norris, 1994) operating on a 20,000-word lexicon of Dutch. For the simulations, all the experimental items were transcribed and used as input to the model. Mid-class phoneme transcriptions (see also Norris, 1994) were added to the Dutch phonemic inventory in SHORTLIST, in order to model the difference between minimal and maximal consonantal changes. Minimal consonantal changes were coded as segments that were defined to be ambiguous between sets of phonemes that differ only in place of articulation. Each mid-class segment was defined so that it matched all phonemes in the class equally well, but mismatched all other phonemes. The maximal consonantal changes were coded as unambiguous phonemic segments, mismatching all other phonemes. The performance of the model on the three different conditions was compared in time slices, one for each phoneme in the input, followed by four time slices of silence. The mean activation functions for the target words and the longer embedding words are shown in Figure 4-3.

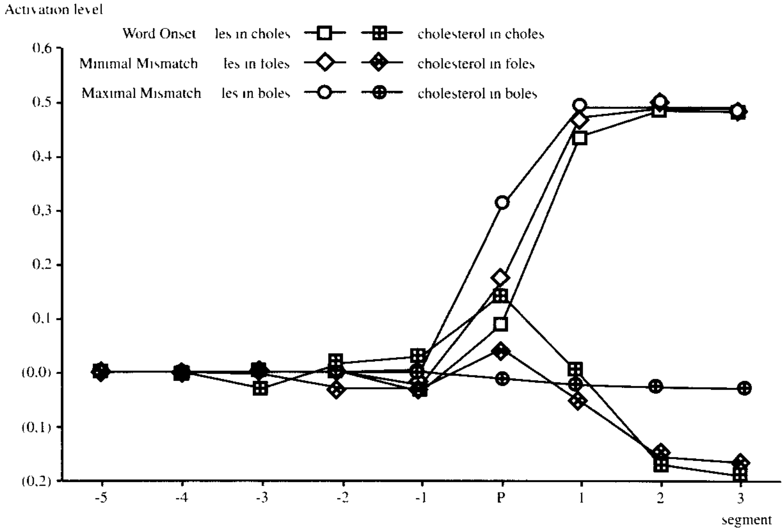

Figure 4-3 *Mean target activation levels for the materials from Experiment 4.1 in SHORTLIST. The model was run on a 20,000-word Dutch lexicon, using the Possible Word Constraint (PWC) described in Norris, McQueen, Cutler & Butterfield (1997). The mean activation functions are shown for the embedded target words and for the competing embedding longer words. Mid-class phoneme transcriptions (see Norris, 1994) were added in order to model the difference between minimal and maximal mismatches. The activations functions are aligned relative to the final phoneme of the embedded word ("P").*

The time course of the predicted competition process can be investigated by looking at the activations of the target words at different time slices. Up to the penultimate phoneme of the experimental item there are no differences in level of activation between items from the three different conditions. At the offset of the target word, targets from the Word Onset condition have a much lower activation than targets from the Maximal Mismatch condition, and targets from the Minimal Mismatch condition seem to be an intermediate case. At the first silence there is hardly any difference in activation left between the Minimal and the Maximal Mismatch condition, but both are still more highly activated than targets from the Word Onset items. At later time slices the three activation functions converge completely.

These simulations confirm the viability of an explanation of the results from the two form priming experiments in terms of the time course of competition. Experiment 4.2 could be tapping into the competition process too early (visual processing starting at about "-2" on Figure 4-3) such that no priming is found because the embedded target is not yet activated. Experiment 4.1 could be tapping too late (visual processing starting at about "P" on Figure 4-3). At this point in time the target word is activated and primes lexical decision in all three conditions, but in this case competition could have acted to rule out the longer candidate in all three conditions before the lexical decision is made. An obvious next experiment would be an identity priming experiment where the visual targets are presented about halfway through the embedded words in the auditory stimulus. However, there is an important problem with such an experiment: what is 'halfway'? The embedded target words were all monosyllabic, but they did not all have the same syllabic structure. The majority of the targets (19) had a Consonant-Vowel-Consonant structure (CVC) and for these items the midpoint could arguably be defined as the middle of the vocalic portion. The other 13 targets, however, had a Consonant-Vowel structure (CV) and it would be hard to establish where exactly 'halfway' is for these items. A different solution would be to just present the visual stimuli halfway in time. For Experiment 4.2 the durations of the embedded target words were measured in order to present the primes at the onset of the embedded word. These durations could be used to present the visual stimuli after half of the duration of the embedded word. However, this approach is even more problematic, since halfway in time is not necessarily halfway in information. For example, halfway in time of the embedded word *les* in *choles* is somewhere during the final fricative [s]. The most serious problem with the identity priming task is that it has failed to detect even a difference between the word-onset and maximal mismatch conditions. It is unlikely that the task is sensitive enough to detect a difference between minimal and maximal mismatch conditions, should there be one.

Given these problems with the identity priming task, it was decided that the word-spotting task would be used to investigate further the effects of mismatching information on lexical access and competition.Word spotting has been shown to be a good methodology to investigate competition effects between lexical candidates (McQueen et al., 1994; Norris et al., 1995). In combination with manipulations of the amount of mismatch it may provide a good measure of the effects of mismatching information on lexical access and competition.

EXPERIMENT 4.3

In Experiment 4.3 participants were asked to listen to a list of bisyllabic nonsense items and to try to spot words embedded at the ends of the stimuli. McQueen et al. (1994) found for English that it is harder to spot a word like *mess* in [dəmɛs] which is the beginning of *domestic*, than in [nəmɛs], which is not the beginning of an English word. This competition effect was utilized in a Dutch word-spotting experiment. It should be harder to spot *les* in [xo:les] than in [bo:les], but what happens in the case of [fo:les], where the initial consonant [f] is much closer to the original [x] than the initial consonant [f]? Will it pattern with the word onset condition [xo:les], with [bo:les] or will it be an intermediate case?

Method

Materials

The materials that were used were the same materials that were used as the auditory stimuli in the two form priming experiments. To obtain a balanced design one experimental triplet, one matched filler triplet and one free filler were added to the stimulus set[2] (see also Appendix 4-A).

The three different conditions, Word Onset, Minimal Mismatch and Maximal Mismatch, for every target and every matched filler were counterbalanced across three different versions of the experiment, so that each version contained a different first syllable for each particular target word and for each matched filler. Each version therefore contained a total of 33 target-bearing bisyllables (11 for each condition), 33 matched fillers with no embedded words (11 for each condition) and a set of 33 free fillers that did not

---

[2] *These items were recorded during the same session as the auditory stimuli that were used in Experiment 4.1 and 4.2.*

form the beginning of any existing word in Dutch and contained no embedded words. Each member of an experimental triplet or matched filler triplet and all free fillers were presented in the same order in all of the three versions.

Subjects

Forty-eight subjects participated in Experiment 4.3 in three groups of 16 subjects for each version. They were all paid for their participation. None of them reported any hearing disorder and none of them had participated in any of the other experiments.

Procedure

The stimuli were presented binaurally via Sennheiser headphones. The presentation of the stimuli, the timing of the manual responses and the collection of the data was again performed by the NESU experimental system. Subjects were tested two at a time in individual sound-attenuated booths. Participants were seated in front of a computer screen, which was used to indicate the beginning and the end of the experiment and the pause between the practice block and the experimental block. Subjects were instructed to listen to a list of nonsense words. They were asked to press a response button with their preferred hand whenever they heard a real word embedded at the end of the bisyllable that was presented, and to say which word they heard. Their answers were monitored on-line for false alarms, that is, for occasions where subjects spotted a word other than the intended target word. Nine practice trials with similar materials were presented before the 99 trials of the experimental block.

Results and discussion

Trials where subjects did not respond at all (9.9%) or when the manual response was not followed by a correct repetition of the intended target word (1.2%) were all treated as errors. Furthermore, responses of less than 200 ms or greater than 1800 ms were also treated as errors. As before, all missing latencies were substituted by the relevant (per subject or per item) mean reaction time per condition. Participants failed to respond correctly in 12% of the target bearing trials. The reaction times were measured from the offset of the target word. The mean word-spotting latencies and errors from the subjects' analysis are presented in Figure 4-4.

ANOVAs were performed on the mean word spotting latencies and errors collapsing across items and subjects respectively. Listeners were on average

119 ms slower in detecting a word that was embedded in a string that formed the beginning of a real word in Dutch. However, there was no difference between the Minimal Mismatch and the Maximal Mismatch conditions. The ANOVAs of the reaction times show that this was a highly reliable effect, $F1(2,90) = 42.92$, $p < .001$, $MSE = 5{,}318$; $F2(2,64)=24.46$, $p < .001$, $MSE = 6{,}673$. Post hoc Tukey HSD tests ($p < .01$) confirmed that the Minimal Mismatch condition and the Maximal Mismatch condition both differed significantly from the Word Onset condition but that they were not significantly different from each other. No reliable differences were observed between the proportions of errors in the different conditions.

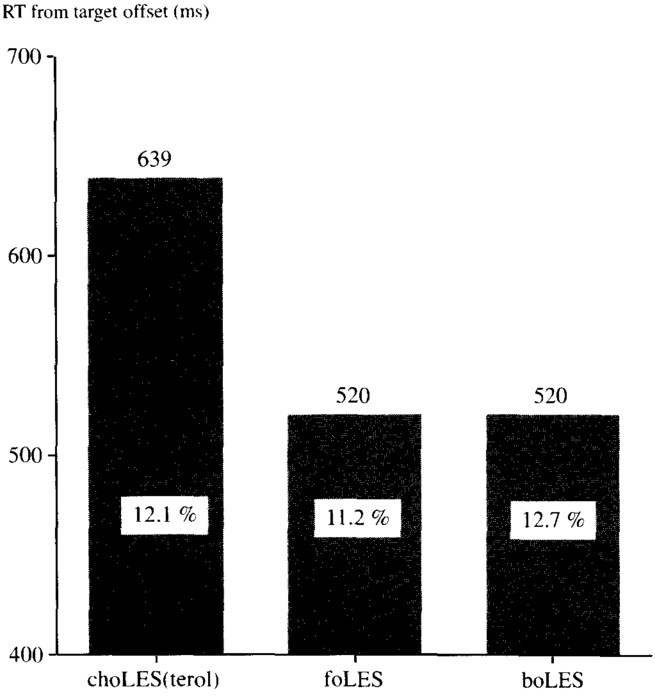

Figure 4-4 *Average word spotting latencies and error rates (inside the columns) for Experiment 4.3. Results are given for each condition: Word Onset (choLES(terol)), Minimal Mismatch (foLES) and Maximal Mismatch (boLES).*

To summarise, the data of this experiment demonstrate a strong competition effect in the detection of words in bisyllabic nonsense forms that can be continued to form longer words. These findings replicate the results from the original study by McQueen et al. (1994). Furthermore, these results suggest that, in this case, any phonemic mismatch is good enough to rule out the longer candidate words, for instance *cholesterol*.

A second word-spotting experiment was conducted to investigate whether differential effects could be found by benefiting the longer candidate words in the competition process. Instead of just recording the first two syllables of the longer word, *choles* from cholesterol, and its derived mismatching counterparts, *foles* and *boles*, one can also record the complete word, *cholesterol*, and its derivatives, *folesterol* and *bolesterol*, and generate bisyllabic stimuli from these items. This method should increase the activation of the longer candidate and disfavour the embedded word. If a differential competition effect could be established between the Minimal and the Maximal mismatch condition by applying this procedure, this would be an interesting finding by itself, but also further confirmation of the utility of the word-spotting task for investigating effects of mismatching information on the competition between misaligned candidates. That is, it would show that the word-spotting task is tapping into the right stage of processing for investigating possible effects of mismatch.

EXPERIMENT 4.4

In Experiment 4.4 listeners were asked to listen to a list of bisyllabic nonsense items that were generated from longer utterances and to try to spot words embedded at the ends of those stimuli.

Method

Materials

The materials were identical to the materials that were used in Experiment 4.3, except for the way in which they were recorded and edited.

Recording

All of the stimuli were recorded in a sound attenuated booth. All of the items were spoken as complete strings of three or more syllables, three times in pseudo-random order by a male native speaker of Dutch. As before, the experimental items and the matched fillers were recorded in triplets, e.g. *cholesterol-folesterol-bolesterol*, in order to minimise the acoustic differences between the three versions, other than the change of the initial consonant. After the recording all items were again digitised at a sampling frequency of 16 kHz. All the materials were measured and the first two syllables of each item were excised from their original context, using the Xwaves/ESPS speech editor. For each triplet the most constant utterance was selected. The target items were again inspected to make sure that all of the resulting target-bearing bisyllables were syllabified correctly between target word and nonsense syllable.

Subjects

Thirty-six subjects participated in Experiment 4.4 in three groups of 12 subjects for each version. None of them had participated in any of the other experiments.

Procedure

The procedure was identical to that of Experiment 4.3.

Results and discussion

Trials where subjects did not respond at all (29%) and false alarms (0.9%) were excluded from the reaction time analyses. Furthermore, responses of less than 200 ms or greater than 1800 ms were also treated as missing data. All missing latencies were substituted by the overall mean reaction time per condition. Participants failed to respond correctly in 33% of the target bearing trials. The reaction times were measured from the offset of the target word. The mean word spotting latencies and errors from the subjects' analysis are presented in Figure 4-5.

As before, ANOVAs were performed on the mean word spotting latencies and errors. Listeners were on average 146 ms slower in detecting a word that was embedded in a string that formed the beginning of a real word in Dutch than in a string that was not a word onset. Furthermore, the listeners were on average 51 ms faster in detecting words in the Maximal Mismatch condition than in the Minimal Mismatch condition. The overall ANOVAs of the reaction times showed a highly reliable effect of the three different conditions (Word Onset, Minimal Mismatch and Maximal Mismatch), $F1(2,66) = 27.03$, $p < .001$, $MSE = 10{,}276$; $F2(2,64)=6.52$, $p < .01$, $MSE = 24{,}427$. The results of post hoc Tukey HSD tests ($p < .01$) revealed an intermediate effect for the Minimal

Mismatch condition. By subjects, the Maximal and the Minimal Mismatch condition differed significantly from the Word Onset condition but not from each other. By items, the only significant difference was between the Maximal Mismatch condition and the Word Onset condition (Item means: Word Onset = 987 ms; Minimal Mismatch = 896 ms; Maximal Mismatch = 851 ms).

Similar analyses were performed for the error data. Listeners made on average 7.5% more errors in the two mismatching conditions than in the Word Onset condition. However, this difference was significant by subjects, $F1(2,66) = 4.29$, $p < .05$, $MSE = .01$, but not by items, $F2(2,64)=2.34$, $p > .1$, $MSE = .02$.

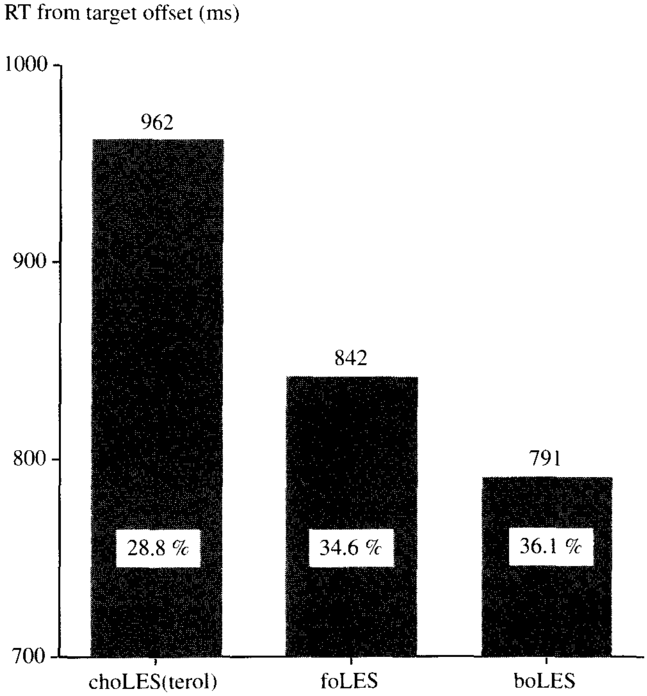

Figure 4-5 *Average word spotting latencies and error rates (inside the columns) for Experiment 4.4. Results are given for each condition: Word Onset (choLES(terol)), Minimal Mismatch (foLES) and Maximal Mismatch (boLES).*

The observed effects could nevertheless be due to some kind of trade-off between speed and accuracy. To further investigate this possibility, six triplets where subjects failed to respond correctly in more than 75 % of the cases to at least one of the three stimuli (e.g. *pa* (dad) in *capa(citeit)-tapa-zapa*) were excluded from the original data set. The mean word spotting latencies and errors from the subjects' analysis on this new data set are presented in Figure 4-6.

Again, ANOVAs were performed on the word spotting latencies and on the proportions of errors. These analyses suggested that the overall effects cannot be explained as a speed-accuracy trade-off, since the effect in RT was robust ($F1(2,66) = 17.59$, $p < .001$, $MSE = 15,676$; $F2(2,52)=7.68$, $p < .01$, $MSE = 24,159$) while the effect in error rate has disappeared completely (all p's $> .5$).

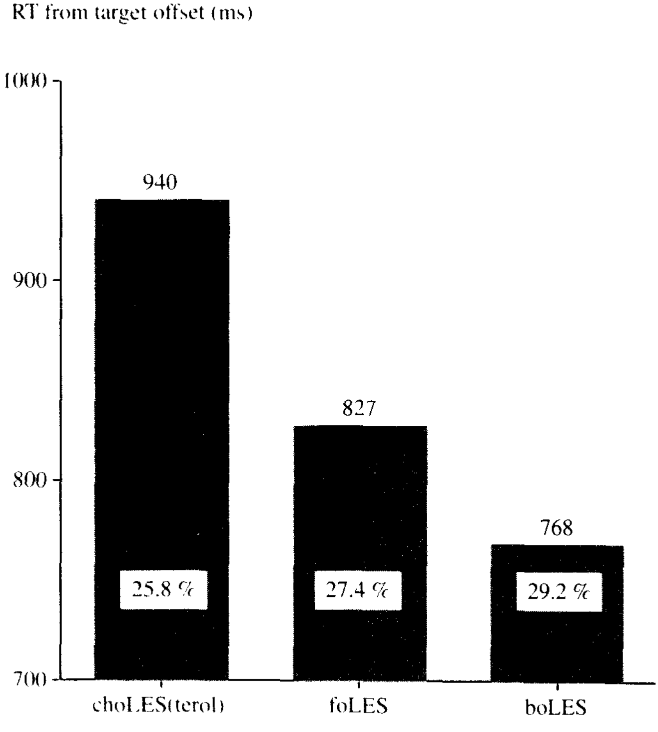

Figure 4-6 *Average word spotting latencies and error rates (inside the columns) for Experiment 4.4 after excluding 6 items where more than 75% of the participants did not respond to at least one of the triplets. Results are given for each condition: Word Onset (choLES(terol)), Minimal Mismatch (foLES) and Maximal Mismatch (boLES).*

The results of post hoc Tukey HSD tests ($p < .01$) again revealed an intermediate effect for the Minimal Mismatch condition. By subjects, the Maximal and the Minimal Mismatch condition again differed significantly from the Word Onset condition but not from each other. By items, the only significant difference was between the Maximal Mismatch condition and the Word Onset condition (Item means: Word Onset = 940 ms; Minimal Mismatch = 827 ms; Maximal Mismatch = 768 ms).

Experiment 4.4 showed the same competition effect as Experiment 4.3. Listeners found it easier to spot words in the two mismatching conditions than in the Word Onset condition. Furthermore, in Experiment 4.4, it was found that it was harder to spot words in the Minimal Mismatch condition than in the Maximal mismatch condition. This suggests that single feature mismatches (e.g. changes in place of articulation) are tolerated more than those mismatching by several features (e.g. changes in place, voice and manner).

GENERAL DISCUSSION

The main aim of this research was to investigate the influence of word initial mismatching information on the competition between lexical candidates. Overall, the results suggest that listeners are intolerant to word initial mismatches at a phonetic level. Two cross-modal form priming experiments were conducted to investigate the effect of mismatching information on lexical access and competition. The data from Experiment 4.1 showed that listeners found it more difficult to make a visual lexical decision to words, for example LES, that were presented after hearing a related fragment prime, for example *les* embedded in a longer utterance like *choles*, *foles* and *boles*, than to the same word after hearing an unrelated fragment prime, for instance *redak*. Furthermore, Experiment 4.1 demonstrated that the lexical decision response to a related visual target that is presented at the offset of the embedded word was equally facilitated in all three of the critical conditions compared to an unrelated baseline control condition. This suggests that, at this point in time, the longer word candidate and the embedded words are activated to a similar extent by all three fragment primes. Experiment 4.2 showed that none of the related fragments primed the lexical decision response when the embedded word was presented visually at the onset of the embedded word in the auditory stimulus. This suggests that the embedded word was not sufficiently activated to prime lexical decision in any of the three conditions. The absence of a competition effect in these priming experiments can be explained in terms of the time-course of the competition process. If visual processing starts at the onset of the embedded word in the auditory stimulus (Experiment 4.2), the embedded word (e.g. *les*) is not activated in any of the three critical conditions, and therefore does not prime visual lexical decision. If visual processing starts at the offset of the embedded word, the embedded word is activated and primes visual lexical decision in all three conditions. The embedding word (e.g. *cholesterol*), however, is activated to a similar degree in all three conditions at this point in time, and does not give rise to a differential effect of competition.

Simulations of the experimental materials using the most recent version of the SHORTLIST model (Norris, 1994) that is adapted to include the Possible Word Constraint (Norris et al., 1997) were run to investigate the viability of this explanation of the results. Experiment 4.2 could be tapping into the competition process too early and Experiment 4.1 could be tapping too late. The option of a third identity priming experiment where the visual targets would be presented about halfway through the embedded words in the auditory stimuli was rejected because it would be impossible to establish an unambiguous informational midpoint for the set of auditory stimuli and because identity priming didn't show a competition effect. A word spotting experiment was conducted instead.

The data from Experiment 4.3 showed that listeners found it more difficult to spot words, for example *les*, that were embedded in a longer utterance that formed the beginning of an existing Dutch word, for example *choles*, than in words that did not form the beginning of a word in Dutch (e.g. *les* in *foles* and *les* in *boles*). Hearing *choles* strongly activates the longer word *cholesterol*, which competes for recognition with the embedded word *les*. This finding replicates the original competition effect that was found by McQueen et al. (1994). No difference was found between minimally and maximally mismatching initial phonemes. The mismatching fragments *foles* and *boles* appear to activate *cholesterol* to a smaller extent than the word onset *choles*, but they do not differentially activate the longer word.

In Experiment 4.4 a second word-spotting experiment was conducted. For this experiment the experimental materials were re-recorded as the full embedding word and its mismatching derivatives. The stimuli were constructed by excising the first two syllables from the longer utterances. The results again showed that listeners found it more difficult to spot words, for example *les*, that were embedded in a longer utterance that formed the beginning of an existing Dutch word, for example *choles* excised from *cholesterol*, than in words that did not form the beginning of a word in Dutch (e.g. *les* in *foles* excised from *folesterol* and *les* in *boles* excised from *bolesterol*). As before, this finding replicates the original competition effect that was found by McQueen et al. (1994). Furthermore, the data indicate that mismatches by only one feature (i.e. place of articulation) are tolerated more than mismatches by several features (i.e. place, voice and manner). Hearing *foles* from *folesterol* activates *cholesterol* to a smaller extent than the word onset *choles* from *cholesterol*, but to a larger extent than *boles* from *bolesterol*.

The word-spotting findings provide converging evidence for the relative sensitivity of the spoken word recognition system to word initial phonetic mismatches. In line with earlier research (Bölte, 1997; Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson, & Zwitserlood, 1989) it was shown in Experiment 4.3 that an input that mismatches by a single feature is not treated as an instance of the original word onset fragment. Furthermore, in line with the results found by Marslen-Wilson and collaborators in their manipulations of phonological distance (Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996), no differential effects were found for the amount of mismatch in Experiment 4.3. On the other hand, in Experiment 4.4 it was shown that single feature mismatches, only differing in place of articulation, were tolerated more than mismatches by several features (place, manner and voice). This is consistent with earlier findings that demonstrated graded effects of mismatch (Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997). Three factors that are relevant to this discrepancy are worth considering before we turn to more general issues.

A first concern is perceptual distance, that is, what constitutes a minimal versus a maximal mismatch? In the present study a minimal change was defined as a change in place of articulation only, and a maximal change was defined as a change in place, voice and manner. Other definitions have been used in previous studies. Marslen-Wilson et al. (1996) tested 'close' mismatches that differed by either only manner of articulation or only place from the original word, versus 'distant' mismatches that differed from the original word by two or three of these broad categories. Connine and collaborators used a more elaborate phonetic feature system (Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997) that was based on the description by Chomsky and Halle (1968). Minimally distinct stimuli were generated by changing two or fewer phonetic features and maximally distinct features by changing at least four features. However, just like Marslen-Wilson et al., they did not distinguish between the actual features that were changed. Confusion studies (e.g. Pols, 1983) have shown that perceptual distance corresponds well with phonological distance measures, like the ones mentioned here. Nevertheless, it should be noted that the actual distance metric could have influenced the results. Some of the changes that are defined as minimal or 'close' might be more mismatching than some of the changes that are defined as maximal or 'distant'. Such inconsistencies would counteract the effects of the amount of mismatch. Keeping control over the actual features that are changed does not rule out completely the possibility of such inconsistencies, but it makes the stimuli within the different distance groups more consistent.

A second important factor in the emergence of graded effects of mismatch could be the listening conditions. If you are talking to a person in a noisy bar, you are more likely to be more tolerant towards the information that reaches the ear. Of course, all of the experiments reported here have investigated spoken word recognition under carefully controlled circumstances. If listeners are required to spot words in lists of isolated well-articulated bisyllabic utterances presented over Sennheiser headphones under ideal noise-free listening conditions, it is hardly surprising that they treat a sound like [f] as an instance of that phoneme and not as a mispronounced [x]. In the final experiment a graded effect of mismatching information was found after a simple manipulation, which changed the listening conditions in a relatively natural way. However, constructing the bisyllables from the recorded complete words and their derivatives might not only have changed the listening conditions, but also the lexical extent of the stimuli.

The third factor, lexical extent, was in this case confounded with the change in listening conditions. Lexical extent was defined by Connine (1994) as the amount of redundant information in a word. In this study the lexical extent of both the embedded word and the entire word onset fragment and its mismatching derivatives are relevant. If the first two syllables are taken from a token of the complete word, they are likely to provide relatively more information in support of the longer candidate word, that is, relative to the support for the embedded word, than if they were produced in isolation. This assumption is supported by duration measurements of the auditory stimuli that were used in Experiment 4.3 and 4.4. The embedding fragments in Experiment 4.3 had an average duration of 774 ms (*SD*=98) compared to a length of 513 ms (*SD*=81) in Experiment 4.4. The average duration of the embedded target words was 446 ms (*SD*=71) in Experiment 4.3 and only 240 ms (*SD*=57) in Experiment 4.4. This means that the lengths of the embedded words were reduced by 46% in Experiment 4.4. However, the lengths of the embedding bisyllables were only reduced by 33%. This relative advantage of the longer candidate words in the competition process not only produces a larger basic competition effect (146 ms in Experiment 4.4 versus 119 ms in Experiment 4.3) but might also be responsible for the emergence of the intermediate effect of Minimal Mismatch. In Experiment 4.4, the amount of matching information in favour of the longer candidate is sufficient to cause an interference effect in spite of word-initial minimally mismatching information, and no interference (or a yet weaker interference effect) with maximally mismatching information[3]. The Shortlist simulation (see Figure 4-3) is probably more representative of Experiment 4.4 than of Experiment 4.3. Due to the recording method, the embedded word (e.g. *les*) dominates more in all conditions in Experiment 4.3, and the longer candidate (e.g. *cholesterol*) probably doesn't get going at all given any initial mismatch, no matter whether it is minimal (e.g. *foles*) or maximal (e.g. *boles*).

A final remark concerns the fact that the present study only investigated the effects of word-initial mismatching information. The beginnings of words, at least intuitively, have a special status in spoken word recognition. Therefore, it might be the case that listeners are more sensitive to initial mismatches than to mismatching information at a later position in a word. In a series of cross-modal priming experiments, Connine, Blasko and Titone (1993) found mismatch effects for medial and final mismatches that were similar to the results for word-initial mismatches. Lexical decision responses to visual targets, for example NEAR, were facilitated by hearing a prime that deviated from a semantic associate in medial position, for instance *logal*. A similar pattern was observed for final mismatches (e.g. *professol*-LECTURE). On the basis of these results they argued that word-initial information is not more important relative to other positions in a word. A similar claim was made by Samuel (1981) on the basis of phoneme restoration effects in different positions.

In conclusion, it has been argued that spoken word recognition is rather intolerant with respect to word-initial phonetic mismatches. Only small amounts of mismatching information are tolerated. The observed mismatch effects in favourable conditions for longer candidate words suggest a graded effect of the amount of mismatch. Mismatch is tolerated more when a segment of an utterance mismatches a candidate word by a single feature than when it

---

[3] *Without a control baseline it is impossible to exclude this possibility.*

differs by several features. The drunk stands a fair chance of getting a smoke by asking for a *shigarette*, but will probably just get a frown if he asks for a *bigarette*.

EFFECTS OF LATE OCCURRING MISMATCHING INFORMATION

CHAPTER 5

ABSTRACT

How tolerant is spoken word recognition with respect to mismatching information at a relatively late position in the word? The results reported in Chapter 4 suggested that listeners are rather intolerant of word-initial phonetic mismatch. The results also suggested that mismatches involving only a single feature are less distinctive than mismatches involving several features. These effects were investigated here for mismatches at a later position in a word. Four word-spotting experiments and two identity priming experiments are reported. The observed effects suggest that spoken word recognition is as sensitive to late phonetic mismatches as it is to word-initial phonetic mismatches. Only small amounts of mismatching information are tolerated, independent of the location of the mismatching information. Again, single feature mismatches are less distinctive than mismatches involving two features.

INTRODUCTION

A drunk who asks for a *shigarette* stands a fair chance of getting a smoke but will probably get no more than a frown if he asks for a *bigarette*. In Chapter 4 it was shown that listeners are rather intolerant of word-initial mismatches. Only small amounts of mismatch are tolerated. At least intuitively, however, the beginnings of words seem to be particularly important for fast recognition. Therefore, it might be the case that listeners are more sensitive to word initial mismatches than to mismatching information at a later position in a word. What happens if a drunk asks for a *cigalette* instead of a *cigarette*?

Position-dependent effects of mismatching information

In a series of mispronunciation detection experiments Cole and collaborators (Cole, 1973; Cole & Jakimik, 1978; 1980) found that subjects were faster in detecting mispronunciations in second syllables of words than in word-initial syllables. This suggests that listeners are more sensitive to later occurring mismatching information. However, in line with Cole and Jakimik (1980), one could argue that this result merely reflects the predictability of the second syllable. In a different mispronunciation detection experiment, Cole, Jakimik and Cooper (1978) found an effect in the detection rates in the opposite direction. A mispronunciation of a word-initial phoneme, for example *made* mispronounced as *nade*, was detected more than twice as often as a mispronounced word-final phoneme, for example *time* mispronounced as *tine*. In line with Marslen-Wilson and Welsh (1978), who observed a similar pattern of results in the number of fluent restorations in a shadowing task, Cole and Jakimik (1980) argued that listeners were less sensitive to phonetic detail following recognition. However, the ends of the mispronounced monosyllables can hardly be considered to follow recognition. For instance, the input [taɪm] does not become unique before the final consonant; [taɪ] can also be continued to form the words *type*, *tile*, *tide*, etcetera. As before, the predictability of the intended word seems to provide a more viable explanation. For example, hearing *nade* in a position where a verb is to be expected, might facilitate the detection of the word-initial mispronunciation of the verbal form *made* in comparison to a less predictable mispronounced noun (e.g. *tine*). This explanation seems to be confirmed by a more recent study. In contrast to the series of experiments by Cole and collaborators, Bölte (1997) found no position-specific effects of mismatching information in a mispronunciation detection experiment using Dutch materials. However, instead of presenting the stimuli in a sentential context, Bölte presented the mispronounced words in isolation. In the absence of contextual information, no differences were found between mismatching information at the onset of a word and mismatching information at the offset of a word. For example, the mispronunciation in *trokodil* from *krokodil* (crocodile) was detected just as fast and accurately as the mispronunciation in *krokodir* from *krokodil*[1].

In a series of ten experiments, Bölte (1997) observed a position-dependent effect of mismatching information in only two of them. In all of the other experiments he did not obtain a differential effect for early versus late mismatching information.

In the first experiment which did show a differential effect, a word correction experiment (Bölte, 1997; see also, van Ooijen, 1996), subjects found it easier (as measured by button-press responses) to recognize the intended word after hearing a near word-initial mismatch (e.g. replying *krokodil* after hearing *trokodil*) than after hearing a distant word-initial mismatch (e.g. *vrokodil*). No differences were found between the manual responses to the base word (e.g. *krokodil*) after hearing a late minimally mismatching stimulus (e.g. *krododir*) and those made after hearing a more distant late mispronunciation (e.g. *krokodit*). However, in a second word correction experiment, where subjects had to name the intended word and speech onset time was measured, this differential effect did not emerge. Therefore, it seems plausible that the absence of an effect for the late mismatches in the first word correction experiment was due to strategic behaviour of the subjects, the different response mode or a combination of the two. In the first experiment, for example, subjects might have pressed the response-button without actually having reconstructed the intended word. A similar point was made by Bölte (1997, p. 74).

In the only other experiment in which Bölte observed a positional effect, a cross-modal form-priming experiment, Bölte observed a graded effect of the amount of mismatch in opposite directions for early and late mismatches. For example, the lexical decision response to a visual presentation of the intended word (e.g. KROKODIL) was facilitated more by hearing a late 'distant' mismatch (e.g. *krokodit*) than by a late 'near' mismatch (e.g. *krokodir*). In

---

1. Bölte (1997) also showed that the employment of a different measurement point by Cole et al. was partially responsible for the difference that Cole et al. had observed and that the onset of the mispronunciation is not the appropriate point from which to measure reaction times.

contrast, initial 'near' mismatches facilitated lexical decisions to the intended word more than initial 'distant' mismatches. This pattern of results can not be explained by any existing model and, as before, seems to reflect some kind of answering strategy for final mismatches. For instance, the high proportion of filler pairs with a late 'distant' mismatch might have suggested that a visual nonword would appear on the screen when a late 'distant' mismatch was heard. On 60 filler trials subjects saw a nonword letter string (e.g. EFPERIMENT) after they had heard a late 'distant' mismatching pseudoword (e.g. *portemodee* from *portemonnee*, wallet), whereas they saw a word on screen after a late distant mismatch on only 10 experimental trials.

In summary, some studies have shown a differential effect for mismatching information at later positions in a word. However, the tasks and the materials that were used in these studies seem to have allowed subjects to develop response strategies.

Position-independent effects of mismatching information

Whalen (1991) investigated the effects of subcategorical word-initial and word-final mismatches in a lexical decision task and a naming task. Whalen found that when an [s] or an [ʃ] was combined with vocalic formant transition information that was appropriate to a different fricative, lexical decision responses were slowed down for both words (e.g. *goose*) and nonwords (e.g. *goosh*). Furthermore, no reliable differences were found between word-final and word-initial fricatives (e.g. *goose/goosh* versus *soup/shoup*) in this respect. Whalen also observed that a large proportion of the nonwords that were used were extremely like real words (e.g. *goosh*), such that the misleading cue, although still subcategorical, was pointing toward a word rather than a nonword. Whalen found that responses to these nonwords were significantly slower than the responses to nonwords in the stable nonword condition (e.g. *giss* and *gish*) that were less like any existing words. These results suggest that, just like word-initial deviant information, word-final mismatching information affects the performance on nonwords via the activation of other possible word candidates. This view is further supported by the findings of Marslen-Wilson and Warren (1994) and McQueen, Norris and Cutler (in press). In both studies reliable word-final mismatch effects were found for nonwords that were created by cross-splicing the final consonant of a nonword with the first portion of a word (e.g. ***smob*** created from ***smog*** and *smob*), but much weaker effects were found for nonwords that were created from other nonwords (e.g. ***smob*** created from ***smod*** and ***smob***). The misleading information, in this case the inappropriate cues in the formant transitions of the vowel preceding the cross-spliced consonants, slowed down responses when the misleading cues were pointing toward a word.

Marslen-Wilson and collaborators (Marslen-Wilson, 1993) investigated the effects of mismatching information at the offset of words in two cross-modal semantic priming experiments. In the first experiment they looked at the sensitivity of the recognition system to deviant information at the offset of monosyllabic words. Three different auditory targets, for example, the word *fleet*, a mismatching nonword counterpart *fleak* (differing from the original word only in place of articulation) or the semantically unrelated control *grace*, were presented auditorily just before presenting a visual target, for example SHIP. They found that participants were faster in making a lexical decision response to SHIP when they had just heard *fleet* in comparison to the unrelated control condition. However, no facilitation was found for the nonword prime *fleak*. Similarly, no facilitation effect was found for mismatching word primes. For instance, hearing *street* facilitated lexical decision to ROAD, but hearing the word counterpart *streak* did not influence the response latency. These results are in line with their previous results for word onsets (Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996). Hearing *task* primed lexical decision to JOB but hearing *dask* did not show any facilitatory effect. Similarly, lexical decision to WOOD was easier after hearing *plank*, but not after hearing *blank*.

In the second experiment Marslen-Wilson (1993) went on to investigate word-final mismatches in disyllables. Again, three different auditory targets, for example, the word *sausage*, a mismatching nonword counterpart *sausin* or the semantically unrelated control *tulip*, were presented auditorily just before presenting a visual target, for example MEAT. Participants were faster in making a lexical decision response to MEAT when they had just heard *sausage* in comparison to the unrelated control condition. However, no facilitation was found for the nonword prime *sausin*. Again, these results are in line with previous results for word onsets with Dutch materials (Marslen-Wilson & Zwitserlood, 1989). For example, hearing *honing* (honey) primed lexical decision to BIJ (bee) but hearing *foning* (nonword) did not show any facilitory effect. Similarly, lexical decision to BIJ was easier after hearing *honing*, but not after hearing *woning* (dwelling).

In six cross-modal semantic priming experiments Connine, Blasko and Titone (1993) found mismatch effects for word-medial and word-final mismatches that were similar to the results that they had previously observed for word-initial mismatches. For initial mismatches they found that semantically related words and derived minimal nonwords primed lexical decisions to visual targets, but no facilitation was found for derived maximal nonwords. For instance, *service* and *zervice* facilitated lexical decision to TENNIS but *gervice* did not. Similarly, it was shown that medial and final minimal mismatches showed comparable levels of priming to word-initial minimal mismatches. For instance, for medial mismatches both *local* and *logal* facilitated lexical decision to NEAR. With the word final mismatches, both *professor* and *professol* facilitated responses to LECTURE. Unfortunately, no derived maximal nonwords were included for both later positions, so these results do not provide us with strong evidence that the word recognition sytem is as sensitive to late mismatching information as to word-initial mismatches. Furthermore, in the absence of the maximal mismatch condition the question whether there is a graded effect of the amount of mismatching information at later positions in a word hasn't been addressed systematically. For word-initial mismatches evidence for graded effects of mismatch was not only obtained using the cross-modal priming paradigm (Connine et al. 1993), but also with a phoneme monitoring task (Connine, Titone, Deelman & Blasko, 1997). They found that the detection of a final consonant is easier as a function of how word-like a stimulus is. For example, detecting [p] in *ketchup* was easier than detecting [p] in *getchup*, detecting [p] in *getchup* was easier than detecting [p] in *retchup*, and detecting [p] in *retchup* was easier than detecting [p] in *gidjup*.

The evidence reported in Chapter 4 also present evidence for graded effects of mismatch in initial position. It was demonstrated using a word-spotting task that detecting *les* (lesson) in *choles*, which is also the beginning of the longer word *cholesterol* (choleterol), was harder than detecting *les* in *foles*, and detecting *les* was harder in *foles* (nonword; single feature mismatch) than in *boles* (nonword; multiple feature mismatch). So far, however, no experiment has directly investigated possible graded effects of later occurring mismatching information.

Although the main point here has been to show that position does not interact with effects of mismatching information, some explanation is required to account for the difference between these two sets of results. Marslen-Wilson and collaborators found strong effects of mismatch, irrespective of the position of the deviant information. On the other hand, Connine and colleagues found some evidence that mismatch is tolerated by the recognition system (also independent of the location of the deviant information). Several factors might be responsible for this discrepancy (see also chapter 4). Task differences, the time-course of recognition (Marslen-Wilson, 1993), the amount of mismatch, and related to this, the perceptual distance criteria, listening conditions and lexical extent (Connine, 1994) all provide possible explanations for the absence versus presence of mismatch effects. For example, Marslen-Wilson tested larger mismatches than Connine (*sausin* versus *professol*) or shorter words (*fleet* versus *professol*).

Taft and Hambly (1986) investigated the influence of the amount of mismatch at later positions in a word using an auditory lexical decision task. They found that the similarity of a nonword to a word affected the reaction times. Responses to nonwords that differed by only one phoneme from a real word were slower than responses to nonwords that deviated by more than one phoneme. For example, subjects were slower in deciding that *rhythlic* (which is only one phoneme different from the word *rhythmic*), was not a word, than in deciding that *rhythlan* (which deviates from *rhythmic* by three phonemes) was not a word. These results provide some evidence for a graded effect of mismatch by showing that it is harder to reject a nonword as a word if it is closer to a word. However, since Taft and Hambly compared responses to nonwords, the possibility remains that these results reflect nonword recognition rather than word recognition. That is, the degree of mismatch might influence the time that listeners need to make the decision that what they heard is not a word, but this does not necessarily involve direct recognition of the base word (e.g. *rhythmic*).

Taken together, the empirical evidence suggests that early and late mismatching information do not have a different status for the word recognition system. The absence versus presence of differential effects of the position of the mismatching information in the signal might be due to differences in the tasks and materials that were used. In Chapter 4 it was emphasized that competition between lexical candidates plays a special role in dealing with mismatching information. A series of two identity priming experiments and two word-spotting experiments investigated the effects of word-initial mismatching information on both lexical access and competition. Overall, the results showed that the word recognition system is rather intolerant of word-initial phonetic mismatches. However, the results also suggested that mismatches involving only a single feature are less distinctive than mismatches involving several features. The aim of the present investigation was therefore to investigate the effects of mismatching information at a later position on both lexical access and competition. The first four experiments used a word-spotting paradigm (Cutler & Norris, 1988); in the last two experiments a cross-modal form priming paradigm (Zwitserlood, 1996) was used. Word-spotting has been shown to be a good methodology to investigate competition effects between lexical candidates (McQueen, Norris & Cutler, 1994; Norris, McQueen & Cutler, 1995). In combination with manipulations of the amount of word-initial mismatch it has also provided a good measure of the effects of mismatching information on lexical access and competition (see Chapter 4).

EXPERIMENT 5.1

In Experiment 5.1 participants were asked to listen to a list of bisyllabic nonsense items and to try to spot embedded words at the beginnings of the stimuli. In Chapter 4 it was found that it was harder to spot a finally embedded target word like *les* (lesson) in [xoːlɛs] which is the beginning of the Dutch word *cholesterol*, than in [foːlɛs] or [boːlɛs], which are not the beginnings of existing Dutch words. This competition effect was utilized in a word-spotting experiment with word-initial embedded targets. It should be harder to spot *ham* (ham) in [hɑmbʉr] which is the beginning of the Dutch word *hamburger* than in [hɑmbʉp], but what happens in the case of [hɑmbʉj] where the final consonant [j] is much closer to the original [r] than the final consonant [p]? Will it pattern with the word onset condition [hɑmbʉr], or with [hɑmbʉp], or will it be an intermediate case?

Method

Materials

Thirty three polysyllabic words with three or more syllables were selected from the Dutch CELEX lexicon (Burnage, 1990). These words all satisfied the following constraints: neither the second syllable nor the first two syllables together formed a real Dutch word, but the first syllable was an existing word in Dutch. For instance, the second syllable and the first two syllables of the Dutch word *hamburger* (hamburger), *bur* and *hambur*, are not existing Dutch words, but the first syllable forms the Dutch word *ham* (ham). For every bisyllabic Word Onset two mismatching versions were created: a Minimal Mismatch condition and a Maximal Mismatch condition. In the minimal change condition the final consonant of the word onset was changed into a consonant that differed only in place of articulation (e.g. *hambuj*); in the maximal change condition the final consonant was substituted by a consonant that differed in place and manner of articulation (e.g. *hambup*)[2].

A set of 66 fillers was also constructed. Half of the fillers, 33, were made to be similar to the target-bearing items: 33 polysyllabic words with three or more syllables were selected from the Dutch CELEX lexicon, but now the first two syllables contained no embedded words. Again, two mismatching versions were created for every bisyllabic word onset. This yielded 33 matched filler triplets. For example, the triplet *sugges-suggef-suggel*, was constructed from the Dutch word *suggestie* (suggestion). These matched fillers were included to discourage listeners from making strategic responses to word onsets. A complete list of the target bearing items and the matched fillers is given in Appendix 5-A. The other 33 fillers did not contain embedded words and did not form the onsets of longer words in Dutch.

The three different conditions, Word Onset, Minimal Mismatch and Maximal Mismatch for every target and every matched filler were counterbalanced across three different versions of the experiment, so that each version contained a different first syllable for each particular target word and for each matched filler. Each version therefore contained a total of 33 target-bearing bisyllables (11 for each condition), 33 matched fillers with no embedded words (11 for each condition) and a set of 33 free fillers that did not form the beginning of any existing word in Dutch and contained no embedded words. The order of presentation of the experimental triplets, matched filler triplets and free fillers was the same in all three versions: the only difference between versions was that in each version a different member of each experimental and matched filler triplet was used.

Simulations

The critical stimuli were first evaluated by running simulations with SHORTLIST (Norris, 1994; Norris, McQueen, Cutler & Butterfield, 1997) operating on a 20,000-word lexicon of Dutch. For the simulations, all the experimental items were transcribed and used as input to the model. Mid-class phoneme transcriptions (see also Norris, 1994) were added to the Dutch phonemic inventory in SHORTLIST in order to model the difference between

---

[2] *In Dutch, only voiceless consonants occur in syllable-final position (Booij, 1995), so it was not possible to manipulate voicing in this position (as in the maximal change conditions in Chapter 4). A change in both place and manner is as large a change as is possible in syllable-final position in Dutch.*

minimal and maximal consonantal changes. Minimal consonantal changes were coded as mid-class inputs that were defined to be ambiguous among sets of phonemes that differed only in place of articulation. The minimal change (mid-class) inputs therefore matched all phonemes in their predefined classes equally well, but mismatched all other phonemes. Maximal consonantal changes were coded as unambiguous phonemic segments, mismatching all other phonemes irrespective of their phonetic class. The performance of the model on the three different conditions was compared in time slices, one for each phoneme in the input, followed by three time slices of silence. The mean activation functions for the target words and the longer embedding words are shown in Figure 5-1.

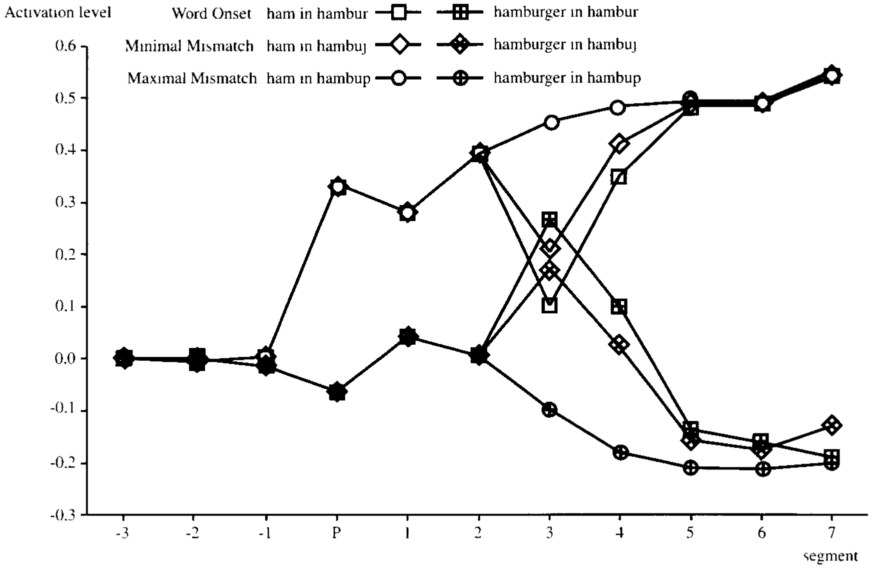

Figure 5-1 *Mean target activation levels for the materials from Experiment 5.1 in SHORTLIST. The model was run on a 20,000-word Dutch lexicon, using the Possible Word Constraint (PWC) described in Norris et al. (1997). The mean activation functions are shown for the embedded target words and for the competing embedding longer words (see text for details). The activation functions are aligned relative to the final phoneme of the embedded word ("P").*

The predicted competition process and its course in time were investigated by looking at the activations of the target words and at the activations of the competing longer word candidates at different time slices. Up to the penultimate phoneme of the experimental item (e.g. after the input [hɑmbʉ]) there are no differences in level of activation for either the targets or the embedding words among the three different conditions. At the offset of the experimental item, targets from the Word Onset condition (e.g. *ham* after the input [hɑmbʉr]) have a much lower activation than targets from the Maximal Mismatch condition (e.g. *ham* after the input [hɑmbʉp]), and targets from the Minimal Mismatch condition (e.g. *ham* after the input [hɑmbʉj]) are an intermediate case. At the first silence there is still a small difference in activation left between the three conditions. At later time slices the three activation functions converge completely. These simulations thus predict that it should be possible to obtain differential effects of late occurring mismatching information, provided that the experimental task is tapping into the right stage of processing.

Recording

All of the stimuli were recorded in a sound damped booth. All of the bisyllabic items were spoken three times in pseudo-random order by a male native speaker of Dutch. The experimental items and the matched fillers were recorded in triplets, e.g. *hambur-hambuj-hambup*, in order to minimize the acoustic differences between the three versions, other than the change of the final consonant. After recording, all items were digitized at a sampling frequency of 16 kHz. All the materials were measured and cut into separate speech files, using the Xwaves/ESPS speech signal processing package. For each triplet the most constant utterance was selected. The target items were inspected to make sure that all target-bearing bisyllables were syllabified correctly between target word and nonsense syllable.

Subjects

Forty-eight subjects participated in three groups of 16 subjects for each version. All the participants were members of the subject pool of the Max Planck Institute for Psycholinguistics. None of the participants reported any hearing problems and they were all native listeners of Dutch. They were paid for their participation.

Procedure

The stimuli were presented binaurally via Sennheiser headphones. The esentation of the stimuli, the timing of the manual responses and the collection of the data was performed by the NESU experimental system. Subjects were tested two at a time in individual sound attenuated booths. Participants were seated in front of a computer screen, which was used to indicate the beginning and the end of the experiment and the pause between the practice block and the experimental block. Subjects were instructed to listen to a list of nonsense words. They were asked to press a response button with their preferred hand whenever they heard a real word embedded at the beginning of the bisyllable that was presented and to say which word they heard. Their answers were monitored on-line for false alarms, that is, for occasions where subjects spotted a word other than the intended target word. Nine practice trials with similar materials were presented before the 99 trials of the experimental block.

Results and discussion

Data analysis, Experiments 5.1-5.3

All manual responses that were followed by an incorrect spoken response were set to zero and treated as missing responses, as were the trials in which subjects failed to respond entirely. Responses of duration less than 200 ms or greater than 1800 ms were also treated as errors. Reaction times were measured from target word offset. The Reaction Times (RTs) and the error rates were submitted to Analyses of Variance (ANOVAs) with both subjects (*F1*) and items (*F2*) as the repeated measure. For the *F1* RT analyses the missing responses of each subject were replaced by the mean of that subject's available responses in that particular condition. For the *F2* RT analyses the missing responses to each target were replaced with the mean of the available responses to that target in that particular condition.

Data analysis and discussion, Experiment 5.1

Six experimental triplets where subjects failed to detect the target word in more than 75% of the cases in at least one of the three stimuli (e.g. *asfal-asfaw-asfam*) were excluded from the analyses, leaving 27 experimental triplets.Subjects did not respond correctly in 37.4% of the critical target-bearing trials (see Figure 5-2 for the mean RTs and error rates from the subjects analysis). They detected words other than the intended target in 4.2% of the trials. In the analysis of the RT data and in the analysis of the errors, there were no significant effects of the three different conditions (all p's > .1).

Subjects found it equally hard to spot words that were embedded at the beginning of an existing Dutch word (e.g. *ham* in *hambur*) and to detect words that were embedded at the beginnings of bisyllables that did not form the onset of a real Dutch word (e.g. *ham* in *hambuj* or *ham* in *hambup*). This is in line with earlier findings by McQueen et al. (1994; Experiment 2) who found that it was as hard to spot the English word *sack* in the Word Onset *sacrif(ice)* as in the Nonword Onset *sackrek* when subjects were only monitoring for word-initial embedded words.

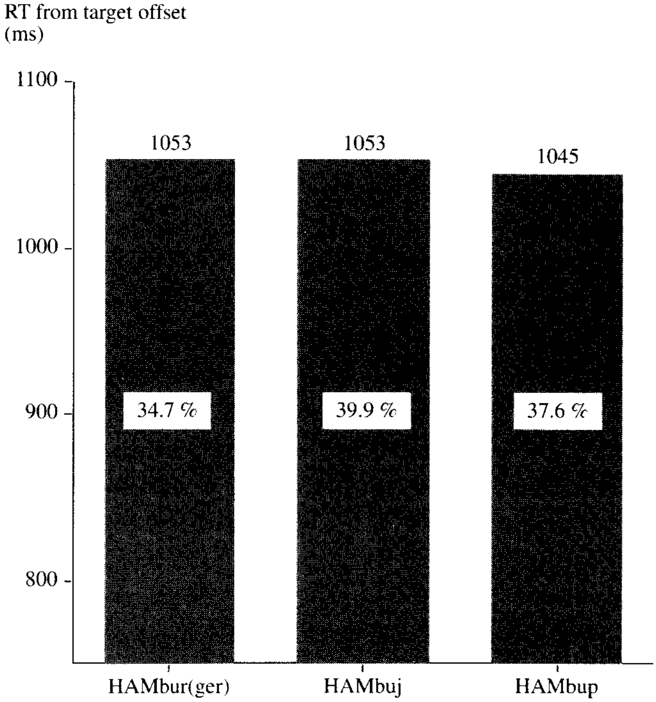

Figure 5-2 *Average word spotting latencies and error rates (inside the columns) for Experiment 1. Results are given for each condition: Word Onset (HAMbur(ger)), Minimal Mismatch (HAMbuj) and Maximal Mismatch (HAMbup)*

The observed effects do not support the initial hypothesis that it should be harder for listeners to detect words that are embedded at the beginning of a bisyllable that can be continued to form a real word. One possible explanation for this absence of an effect of competition could be that subjects had already detected the embedded word (e.g. *ham*) before the longer word candidate (e.g. *hamburger*) was sufficiently activated to interfere with this detection. Alternatively, it might also be the case that the longer candidate word was activated to a similar degree in all three conditions at the time that subjects detected the embedded word. One cannot rule out either of these possibilities in the absence of statistically significant differences between the Word Onset, the Minimal Mismatch and the Maximal Mismatch condition. The simulation data (see Figure 5-1) show that no differential effects are to be expected both when the experiment is tapping too early into the recognition process (before the final phoneme of the embedded word "P" in Figure 5-1) and when the experiment is tapping too late (after the first silence "5" in Figure 5-1).

In a previous word-spotting experiment that investigated word-initial mismatches (Chapter 4) a different recording method was used. Instead of just recording the first two syllables of the longer word, *choles* from *cholesterol*, and its derived mismatching counterparts, *foles* and *boles*, the complete word, *cholesterol*, and its derivatives, *folesterol* and *bolesterol*, were recorded and bisyllabic stimuli were generated by truncating these items. It was demonstrated that this method increases the activation of the longer candidate and disfavours the embedded word, producing a larger competition effect at a later point in time. Therefore, it was hypothesized that applying this recording procedure could result in the emergence of a differential competition effect here (particularly if Experiment 5.1 had tapped too early into the competition process). In Experiment 5.2, therefore, this method was applied to the construction of the experimental materials. Instead of recording only the first two syllables of the longer word, for instance *hambur* from *hamburger*, and its derived mismatching counterparts, *hambuj* and *hambup*, the speaker recorded the complete word, *hamburger*, and its derivatives, *hambujger* and *hambupger*, and bisyllabic stimuli were excised from these items.

EXPERIMENT 5.2

In Experiment 5.2 listeners were asked to listen to a list of bisyllabic nonsense items that were extracted from longer utterances and to try to spot words at the beginnings of those stimuli. As before, it was hypothesized that it should be harder for listeners to detect words that are embedded at the beginning of a bisyllable that can be continued to form a real word than to detect words that are embedded in nonword onsets.

Method

Materials

The materials were identical to the materials that were used in Experiment 5.1, except for the way in which they were recorded and prepared.

Recording

All of the stimuli were recorded in a sound-attenuated booth. All of the items were produced as complete strings of three or more syllables, three times in pseudo-random order by a male native speaker of Dutch. As in Experiment 5.1, the experimental items and the matched fillers were recorded in triplets, e.g. *hamburger-hambujger-hambupger*, in order to minimise the acoustic differences between the three versions, other than the change of the final consonant. After the recording all items were again digitised at a sampling frequency of 16 kHz. All the materials were measured and the first two syllables of each item were excised from their original context, using the Xwaves/ESPS speech editor. For each triplet the most constant utterance was selected. The target items were again inspected to make sure that all of the resulting target-bearing bisyllables were syllabified correctly between target word and nonsense syllable.

Subjects

Thirty-six subjects participated in three groups of 12 subjects for each version. None of them had participated in Experiment 5.1.

Procedure

The procedure was identical to that of Experiment 5.1.

Results and discussion

Overall, participants failed to respond correctly in 47% of the target bearing trials (before items were excluded). Nine experimental triplets where subjects failed to detect the target word in more than 75% of the cases in at least one of the three stimuli were excluded from the analysis, leaving 24 experimental triplets.The mean RTs and errors from the subjects' analysis are presented in Figure 5-3.

Listeners were equally fast in detecting a word that was embedded in a string that formed the beginning of a real word in Dutch as in a string that was not a word onset. Furthermore, the listeners were equally fast in detecting words in the Maximal Mismatch condition as in the Minimal Mismatch condition. The overall ANOVAs of the reaction times showed no differences between the three conditions (*F1* and *F2*, $p > .1$). However, as can be seen in Figure 5-3, the error rates were very high even after excluding the nine triplets with over 75% missing responses in at least one of the three conditions (47% of the remaining 24 targets were missed in the Maximal Mismatch condition). Because of this large amount of missing data for the RT analysis, the error rates might provide a more reliable dependent measure.

Listeners made on average 13% fewer errors in detecting a word that was embedded in a string that forms the beginning of a real word in Dutch than in a string where the final consonant maximally mismatches this word. Furthermore, they made even fewer errors when the targets were embedded in a string where the final consonant minimally mismatches the word. The overall ANOVAs of the error rates showed a highly significant effect of the three different conditions (Word Onset, Minimal Mismatch and Maximal Mismatch), $F1(2,66) = 10.36$, $p < .001$, $MSE = .03$; $F2(2,46)= 9.14$, $p < .001$, $MSE = .02$. Post-hoc Tukey HSD tests by subjects and by items ($p < .05$) indicated that the Minimal Mismatch condition and the Word Onset condition both differed significantly from the Maximal Mismatch condition but that they were not significantly different from each other (Item means, proportion of errors: Word Onset = 34.7%; Minimal Mismatch = 30.2%; Maximal Mismatch = 46.9%).

Inspection of the individual subject means revealed a mixed pattern of results. Some subjects showed an effect in the predicted direction, and made more errors in the Word Onset Condition than in the Maximal Mismatch condition. However, more than half of the subjects (22) made fewer errors in the Word Onset condition than in the Maximal Mismatch condition. A possible explanation for this effect in the direction opposite to that predicted could be that these participants were strategically responding to word onsets. Taken together, these two groups of responses yield a pattern of results where the error rate is lowest in the Minimal Mismatch condition. In contrast to previous experiments that investigated word-initial mismatching information (Chapter 4) there was a bias in the set of materials that were used in the present experiments since all of the target-bearing items and all of the matched fillers are also word onsets up to the final consonants. In Experiment 5.2 this bias in the materials may have been more apparent than in Experiment 5.1 because the recording method favoured the longer candidate words. There are several ways in which the subjects could have exploited this bias.

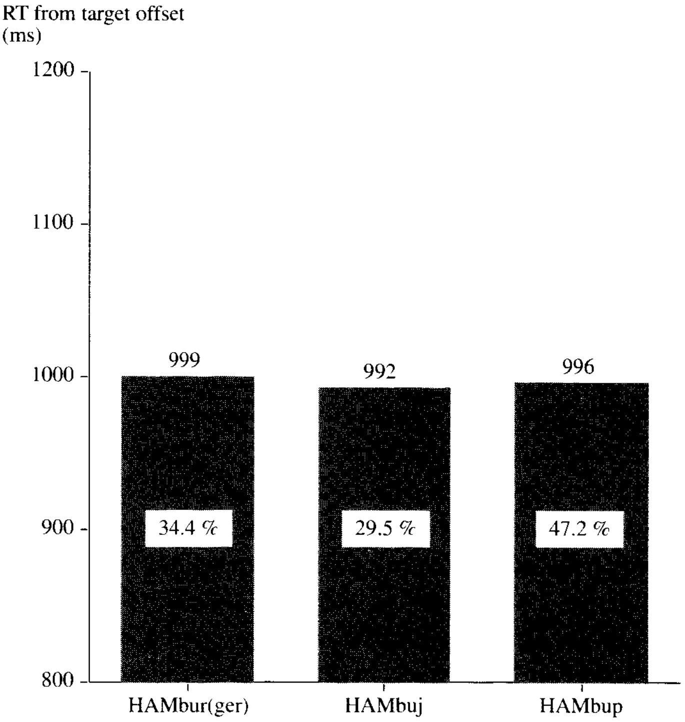

Figure 5-3 *Average word spotting latencies and error rates (inside the columns) for Experiment 5.2. Results are given for each condition: Word Onset (HAMbur(ger)), Minimal Mismatch (HAMbuj) and Maximal Mismatch (HAMbup).*

A first possibility is that the participants expected embedded words whenever they heard a word onset and might then have made more errors when the bisyllabic stimulus turned out to be a nonword onset at the final consonant. In addition, the subjects could even have recognized the embedded target words (e.g. *ham*) via the recognition of the embedding words (e.g. *hamburger*). When using such a strategy, participants would clearly find it harder to recognise the embedding word in the two mismatching conditions. Recognition of the targets via the embedding words might still reflect auditory processing, but in a language like Dutch one cannot rule out the possibility that subjects perform some kind of visual word spotting with their mind's eye on an orthographic representation of the longer candidate word[3]. The Dutch language uses an alphabetic writing system with a very transparent correspondence between letters and sounds, which might allow the participants to apply an orthographic strategy. Although the Dutch spelling system is not ideal in the sense that it deviates from a one-to-one correspondence (Nunn, 1998), the correspondence has been shown to be good enough to influence performance on another auditory task. Dijkstra, Roelofs, and Fieuws (1995) demonstrated that phoneme monitoring latencies were affected by the orthographic representation of Dutch words. They found that it was easier to detect a phoneme in a Dutch word with a primary spelling (e.g. [k] in *perzik*) than in a Dutch word with a secondary spelling (e.g. [k] in *plastic*). However, similar effects have been reported for phoneme monitoring (Frauenfelder, Segui, & Dijkstra, 1990) and for other auditory tasks, such as syllable detection (Dupoux & Mehler, 1992) in French, a language with a much more irregular mapping between letters and sounds than Dutch. This suggests that, besides the characteristics of the spelling system, specific task demands can induce the use of orthographic information, even for languages with a less regular spelling. A similar claim was made by Cutler, Treiman and van Ooijen (1998), who obtained orthographic inconsistency effects in English, which is notorious for its irregular spelling. They suggested that effects of orthographic inconsistency may be artefacts of the use of salient orthographic manipulations that draw the listeners' attention to spelling information. In summary, together with earlier research, the present findings suggest that listeners whose native language has a good correspondence between letters and sounds can use the spelling of words to accomplish an auditory task like word-spotting in that language. Whether this would carry over to similar experiments in languages with a more irregular orthography, like English and French, remains to be investigated.

To investigate further the strategic behaviour of the participants, an additional word-spotting experiment was conducted in which extra filler trials were presented. On these trials the subjects were presented with nonword onsets, that is, bisyllabic stimuli that did not form the beginnings of real words in Dutch. These fillers were added to change the proportion of word onsets in the experiment. It was predicted that adding fillers which did not form the beginnings of any real words would make the 'word onset'-bias in the stimulus

---

3. One of the participants actually reported during debriefing that she had used such a strategy. Further evidence for an orthographic strategy comes from listeners' errors. Seventeen percent of the subjects who heard justit, *[*jʉsti:t*], for example falsely detected the word* jus *(gravy; [*ʒy:*]), which is an orthographic but not a phonological embedding.*

set less apparent, and thus would discourage listeners from using an orthographic strategy.

EXPERIMENT 5.3

In Experiment 5.3 listeners were asked to listen to a list of bisyllabic nonsense items that were generated from longer utterances and to try to spot words at the beginnings of those stimuli. They were presented with stimuli that did not form the beginning of existing Dutch words on two-thirds of the trials (rather than on only one-third of the trials, as in Experiments 5.1 and 5.2).

Method

Materials

The materials were identical to the materials that were used in Experiment 5.2, except for the extra filler items. Sixty-six additional nonword onset fillers were taken from the stimulus materials of a previous word-spotting experiment that investigated word-initial mismatches (Chapter 4; Experiment 4.4). As before, the three different conditions, Word Onset, Minimal Mismatch and Maximal Mismatch for every target and every matched filler were counterbalanced across three different versions of the experiment, so that each version contained a different first syllable for each particular target word and for each matched filler. Each version therefore contained a total of 33 target-bearing bisyllables (11 for each condition), 33 matched fillers with no embedded words (11 for each condition) and a set of 99 free fillers without initially embedded words.

Recording

All of the stimuli, including the additional fillers, were recorded during the same recording session. The experimental items were the same as in Experiment 5.2.

Subjects

Thirty-six subjects participated in three groups of 12 subjects for each version. None of them had participated in Experiment 5.1 or 5.2.

Procedure

The procedure was identical to that used in Experiments 5.1 and 5.2.

Results and discussion

Overall, participants failed to respond correctly in 49.8% of the target bearing trials. Nine experimental triplets where subjects failed to detect the target word in more than 75% of the cases in at least one of the three stimuli were excluded from the analyses, leaving 24 experimental triplets.The mean word-spotting latencies and errors from the subjects' analysis are presented in Figure 5-4.

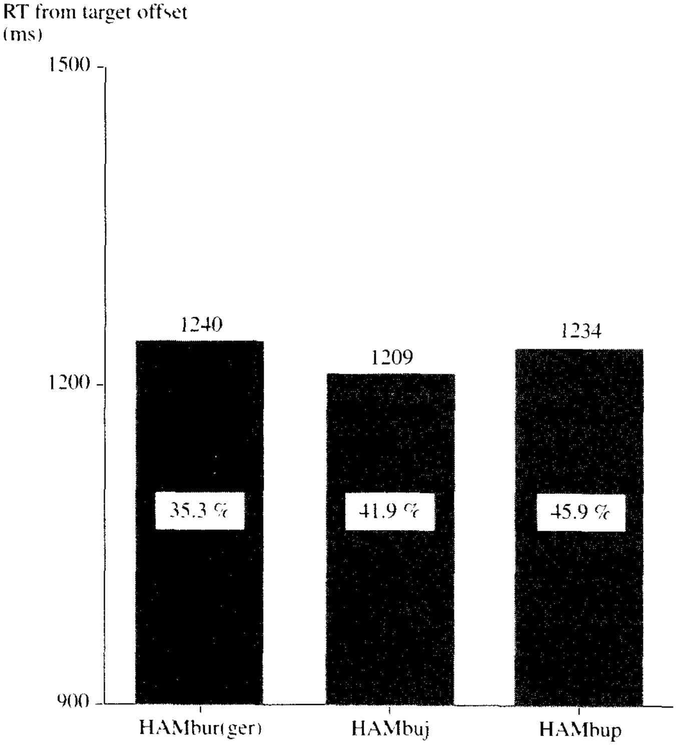

Figure 5-4 *Average word spotting latencies and error rates (inside the columns) for Experiment 5.3. Results are given for each condition: Word Onset (HAMbur(ger)), Minimal Mismatch (HAMbuj) and Maximal Mismatch (HAMbup).*

Listeners were equally fast in detecting a word that was embedded in a string that formed the beginning of a real word in Dutch as in a string that was not a word onset. Furthermore, the listeners were equally fast in detecting words in the Maximal Mismatch condition as in the Minimal Mismatch condition. The overall RT ANOVAs showed no differences between the three conditions ($F1$ and $F2$, $p > .1$). However, as can be seen in Figure 5-4, the error rates were very high even after excluding the nine triplets with over 75% missing responses in at least one of the three conditions (46% of the remaining 24 targets was missed in the Maximal Mismatch condition). Therefore, as in Experiment 5.2, the more reliable dependent measure is provided by the error rates.

Listeners made on average 11% fewer errors in detecting a word in the Word Onset condition than in the Maximal Mismatch condition. Furthermore, they made 7% fewer errors in the Word Onset condition than in the Minimal Mismatch condition. The overall ANOVAs of the error rates showed a marginally significant effect of the three different conditions (Word Onset, Minimal Mismatch and Maximal Mismatch), $F1(2,66) = 4.60$, $p < .05$, $MSE = .02$; $F2(2,46)= 2.04$, $p < .1$, $MSE = .03$. Post-hoc Tukey HSD tests by subjects ($p < .05$) indicated that the only significant difference was between the Maximal Mismatch condition and the Word Onset condition. Similar analyses by items showed no reliable differences between any of the three conditions (Item means, proportion of errors: Word Onset = 36.5%; Minimal Mismatch = 42.7%; Maximal Mismatch = 45.8%)

These results suggest that adding nonword fillers did not have the intended effect on the strategic behaviour of the participants. Although the differences in the error rate between the three conditions were no longer completely reliable when compared to the differences in Experiment 5.2, they still showed a trend in the opposite direction from what was originally predicted.

In other words, it should have been harder to detect a word that was embedded at the beginning of a longer word in Dutch than to spot a word that was embedded in a nonword onset. However, the results from Experiments 5.2 and 5.3 indicate that it was easier to detect words that were embedded in word onsets. The word onset bias that was present in the critical stimuli sets in Experiments 5.2 and 5.3 provides a viable explanation for these reversed patterns of results. Whenever a response was required, that is, whenever an auditory stimulus started with an embedded word, the auditory stimulus formed the beginning of an existing word in Dutch at least up to the final consonant. For example, *ham* is embedded in the word onset [hɑmbʉ] in all three conditions[4].

In a second attempt to make the experiment less susceptible to response strategies that were, most likely, induced by the word onset bias in the materials, a different approach was taken. It was predicted that running a word-spotting experiment with target words in both stimulus-initial and stimulus-final position would prevent the participants from recognising the embedded target words via the recognition of the embedding words. Such a strategy would be much less effective given a final target. For example, it should be harder to recognize *les* via recognition of *cholesterol* than to recognize *ham* via *hamburger*. Furthermore, McQueen et al. (1994) found no effects of competition when subjects were only monitoring for word-initially embedded targets (McQueen et al., 1994; Experiment 2), but they did find such an effect in the error rates for initial targets when subjects were also monitoring for final targets (McQueen et al., 1994; Experiment 1). They argued that specification of the target location allowed the listeners to respond fast enough for the competition effect to go undetected. In their first experiment, however, in which targets could appear in either location, the participants responded late enough for small differential competition effects to have emerged.

EXPERIMENT 5.4

In Experiment 5.4 the experimental items from Experiment 5.2 were mixed with the experimental items from a previous study (Chapter 4, Experiment 4.4)

---

[4] *The same bias was present in the materials from Experiment 5.1. Inspection of the individual subject data shows that a much smaller proportion of the participants in Experiment 5.1 than in Experiments 5.2 and 5.3 found it easier to detect target words that were embedded in word onsets. Recording the stimuli as bisyllabic utterances should have made the word onset bias less apparent to the participants in comparison to when the items were constructed from recordings of the complete longer word candidates. Nevertheless, a similar control experiment to Experiment 5.3 was conducted using materials recorded as in Experiment 5.1, because even a small bias effect could have acted to obscure an effect of competition. The results of this experiment were similar to the results from Experiment 5.1. Listeners found it equally hard to detect a word when it was embedded in a word onset and when it was embedded in a bisyllable that mismatched the word onset at the final consonant. However, since adding fillers was not very effective in Experiment 5.3, this control experiment did not provide any strong evidence that the absence of a differential effect of competition in Experiment 5.1 was caused by a small bias effect cancelling out a small effect of competition.*

that investigated word-initial mismatches. The participants were now asked to listen to a list of bisyllabic nonsense items and to try to spot words embedded both at the beginnings and at the ends of the stimuli.

Method

Materials

The experimental materials for Experiment 5.4 were taken from the stimulus sets of Experiment 5.2 and a previous word-spotting experiment on word-initial mismatches (Chapter 4; Experiment 4.4). From the original sets of thirty-three target-bearing items, 18 items with word-initial embedded words (*hambur-hambuj-hambup*) and 18 items with word-final embedded words (*choles-foles-boles*) were selected. The following selection criteria were used. Items where listeners had detected the target word in most cases in the original single-position word-spotting experiments were included under the constraint that a word could only appear once as the target in the present experiment. In cases where the same target appeared in both subsets of the 18 best items (i.e. with the lowest error rates) the item with the lowest error rate was selected and, for the counterpart with the same embedded word in the other position, the next best available item was included as an alternative. The target bearing items are given in Appendices 5-A and 5-B.

Seventy-two filler stimuli were also included. Thirty-six matched filler triplets, 18 with a stimulus-initial mismatch and 18 with a stimulus-final mismatch, were selected at random from the two original matched fillers sets. The other 36 fillers were selected at random from the two original free filler sets.

The triplets for every target and every matched filler were again counterbalanced across three different versions of the experiment, so that a particular original or derived word onset appeared only once in each version. Each version therefore contained a total of 36 target-bearing bisyllables (18 initial, 18 final, 6 for each condition), 36 matched fillers with no embedded words (18 initial, 18 final, 6 for each condition) and a set of 36 free fillers that did not form the beginning of any existing word in Dutch and contained no embedded words.

Subjects

Thirty-six subjects participated in three groups of 12 subjects for each version. Persons that participated in any of the other experiments were excluded. None of the participants reported any hearing problems. They were paid for their participation.

Recording

All of the stimuli were recorded during the same recording session, where the complete embedding words were produced by a male native speaker of Dutch. The bisyllabic stimuli were constructed from these longer utterances, as in Experiment 5.2.

Procedure

The procedure was identical to that of Experiment 5.1 except for the instructions. As in Experiment 5.1, subjects were instructed to listen to a list of nonwords. But now they were asked to press a response button with their preferred hand whenever they heard a real word embedded at the beginning or at the end of the bisyllabic stimulus that was presented. Twelve practice trials with similar materials were presented before the 108 trials of the main experiment.

Results and discussion

Not surprisingly, participants found this experiment much harder than the previous single-position word-spotting experiments. Maintaining the same absolute cutoff value (1800 ms) as for the previous analyses would imply excluding responses well within a single standard deviation of the mean for several subjects and items. Therefore, it seemed inappropriate to use this cutoff value. A more liberal measure was used instead. Responses deviating more than two standard deviations from the relevant mean (that is, the mean latencies and standard deviations per subject for the subjects analyses and the mean latencies and standard deviations per item for the items analysis) were treated as errors.

As before, all manual responses that were followed by the wrong verbal response (6.9%) and trials where participants failed to respond entirely (37%) were set to zero and treated as missing responses.Two experimental triplets and six subjects had to be excluded from the analyses because no responses were available for one or more of the experimental conditions. The mean word-spotting latencies and errors from the subjects' analysis are presented in Figure 5-5.

ANOVAs were performed on the mean word-spotting latencies and errors. Listeners were on average 155 ms slower in detecting a word that was embedded in a string that formed the beginning of a real word in Dutch than in a string that was not a word onset. Furthermore, the listeners were on average 55 ms faster in detecting words in the Maximal Mismatch condition than in the Minimal Mismatch condition.

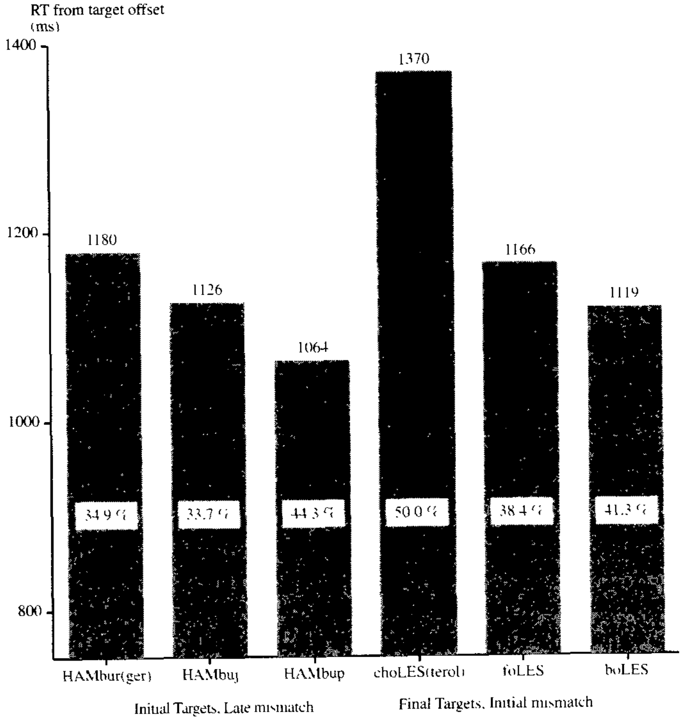

Figure 5-5 *Average word spotting latencies and error rates (inside the columns) for Experiment 5.4. Results are given for all three conditions in two different positions: Initial Targets: Word Onset (HAMbur(ger)), Minimal Mismatch (HAMbuj) and Maximal Mismatch (HAMbup) and Final Targets: Word Onset (choLES(terol)), Minimal Mismatch (foLES) and Maximal Mismatch (boLES).*

The overall RT ANOVAs showed a highly reliable effect of the three different mismatch conditions (Word Onset, Minimal Mismatch and Maximal Mismatch), $F1(2,54) = 11.73$, $p < .001$, $MSE = 45,412$; $F2(2,64) = 8.08$, $p < .01$, *MSE* = 53,847. The results of post hoc Tukey HSD tests ($p < .01$) revealed an intermediate effect for the Minimal Mismatch condition. By subjects, the Maximal and the Minimal Mismatch condition differed significantly from the Word Onset condition but not from each other. By items, the only significant difference was between the Maximal Mismatch condition and the Word Onset condition (Item means: Word Onset = 1344 ms; Minimal Mismatch = 1194 ms; Maximal Mismatch = 1122 ms).

On average, listeners were 95 ms faster in spotting initial targets than in spotting final targets. However, the factor position (Initial Target, Final Target) only reached significance by subjects ($F1(1,27) = 5.23$, $p < .05$, $MSE = 45{,}412$) and not by items ($p > .1$). The interaction of position and mismatch never reached significance (all *p*'s > .1), suggesting that the effects of mismatch for the stimulus-initial mismatches (Final Targets) were equivalent to those for the stimulus-final mismatches (Initial Targets).

Similar analyses were performed for the error data. Across target position there was an effect of mismatch: listeners made on average 6% fewer errors in the Minimal Mismatch condition than in the Word Onset and the Maximal Mismatch condition. This effect was significant by subjects, $F1(2,54) = 3.21$, $p < .05$, $MSE = .03$, but not by items, $F2(2,64) = 1.23$, $p > .1$, $MSE = .03$, and furthermore it was not equivalent across position. On initial targets listeners made on average 10% fewer errors in the Word Onset and the Minimal Mismatch condition than in the Maximal Mismatch condition. For final targets, the effect was in the opposite direction; listeners made on average 10% more errors in the Word Onset condition than in the two mismatching conditions. This pattern of results was confirmed by an interaction of position and mismatch. It was only significant by subjects: $F1(2,54) = 3.37$, $p < .05$, $MSE = .04$, $F2(2,64)=2.22$, $p > .1$, $MSE = .03$. Posthoc pairwise comparisons between the three pairs of mismatch conditions within target position using *t*-tests with a Bonferroni adjusted alpha level ($p < .016$) were conducted to further inspect the interaction. These analyses showed no reliable differences between any two conditions for either the initial or the final targets. Nevertheless, these effects in the error rates might still reflect some specific response strategy to initially embedded targets.

In summary, in Experiment 5.4 a similar pattern of RT results was found for word-initial mismatches combined with stimulus-final embeddings as for mismatches at a later position in a word combined with stimulus-initial embedded words. Experiment 5.4 showed the same competition effect for word-initial mismatches as the original experiment with single-position word-spotting (Chapter 4; Experiment 4.4). Listeners found it easier to spot words in the two mismatching conditions than in the Word Onset condition. As before, it was also found that it was harder to spot words in the Minimal Mismatch condition than in the Maximal mismatch condition. This confirms that single-feature word-initial mismatches (e.g. changes in place of articulation) are tolerated more than those mismatching by several features (e.g. changes in place, voice and manner). Furthermore, Experiment 5.4 extended this finding to word-final mismatches. Listeners found it easier to spot words in the two mismatching conditions than in the Word Onset condition. This is in line with the previous results by McQueen et al. (1994) who only found an effect of competition for initially embedded words, for instance *sack* in *sacrif(ice)* versus *sack* in *sackrek*, when subjects were listening for targets in both positions (McQueen et al., 1994; Experiment 1). Furthermore, it was also found that it was harder to spot words in the Minimal Mismatch condition than in the Maximal Mismatch condition. This confirms that, like early mismatches, single feature late mismatches (e.g. changes in place of articulation) are tolerated more than those mismatching by two features (e.g. changes in place and manner).

However, the pattern of results that was observed in the error data still raises some doubts about whether the participants were perhaps applying a specific response strategy for the initial targets. Therefore, an identity priming task was used to further investigate the effects of late mismatching information on lexical access and competition. Identity priming, like word spotting, has been shown to be a good methodology to investigate competition effects between lexical candidates (Vroomen & de Gelder, 1995; but see Chapter 4). In combination with manipulations of the amount of mismatch it may provide an additional measure of the effects of late mismatching information.

In cross-modal form priming experiments, participants are presented with a visual target right after hearing an auditory prime (Zwitserlood, 1996). A response (lexical decision in most studies) is required to the visual targets. The auditory primes are speech fragments, spoken words or longer utterances that are structurally related to the visual target in the critical conditions. When visual target and auditory prime are related, the auditory prime is intended to facilitate or inhibit the processing of the visual target. This effect is measured with respect to a control condition in which the auditory and the visual stimuli are not related. A major advantage of this paradigm is the possibility that one can obscure the relation between the critical auditory stimuli by adding appropriate filler prime-target pairs. Another advantage is the possibility to present a visual target at various times relative to the presentation of the visual target. This allows one to trace the time course of recognition.

In the following section two form priming experiments are reported in which the participants were asked to perform a different task on a significant proportion of the trials. These catch-trials were added to make sure that the participants were attending to the auditory stimuli (see Chapter 4, Experiments 4.1 and 4.2). As before, in the word spotting experiments, the main aim in these identity priming experiments was to investigate whether competition effects, modulated by the amount of late mismatching information, could be found at different stages in the recognition process. The first priming experiment investigated whether competition effects arise at a relatively early stage of processing.

EXPERIMENT 5.5

In Experiment 5.5 participants were asked to listen to a list of bisyllabic nonsense items. While the listeners were hearing the auditory stimulus two things could happen. In the first case a letter string, for example HAM, could be presented on a computer screen at the offset of *ham* in *hambur*. In this case the participants were asked to decide whether the letter string was an existing Dutch word. In the second case, a warning signal could be presented on the screen at the end of the first syllable of the auditory stimulus followed by the visual presentation of a letter string, for example ZUIF after hearing *zuifkaaf*. In this case the participants were asked to decide whether the letter string was an adequate description of the beginning of the nonsense word that they had just heard. Participants were forced to attend to the auditory stimuli in order to be able to perform well on these catch-trials.

One would predict, in line with the results obtained by Vroomen and de Gelder (1995), that visual lexical decisions on target words that are preceded by auditory stimuli that have those words embedded in them will be less facilitated when the auditory stimuli also form the onsets of longer words in Dutch. Competition with the longer candidate should lower the activation of the word embedded in the auditory stimulus, and thereby reduce the priming effect on the visual lexical decision on the embedded word. For instance, it should be harder to make a lexical decision to HAM after hearing [hambʉr], which is the beginning of *hamburger*, than after hearing [hambʉp]. However, what happens in the case of [hɑmbʉj], where the final consonant [j] is much closer to the original [r] than the final consonant [p]? Will it pattern with the word onset condition [hɑmbʉr], with [hɑmbʉp] or will it be an intermediate case? The results from Experiment 5.4 suggest that [hambup] will mismatch with *hamburger* to block activation of that word, so there will be little or no competition, and lexical decisions to HAM should therefore be relatively easy. The results of Experiment 5.4 also suggest that [hɑmbʉj] may partially activate *hamburger*, and thus that lexical decisions to HAM should be slower than with [hambup] but faster than with [hɑmbʉr].

Method

Materials

The materials from Experiment 5.1 were used as the auditory primes in the present experiment. This choice, of items which had been recorded as bisyllables, was motivated by the idea that the word onset bias would be less apparent in such items than in items which had been recorded as complete words. For all of the auditory stimuli the offset of the first syllable was determined by visual and auditory inspection of the waveform. For this purpose the Xwaves speech editor was used. To obtain a balanced design one experimental triplet, one matched filler triplet and one free filler were removed from the original stimulus set of 33 from Experiment 5.1. The experimental triplet with the highest error rate in Experiment 5.1 was excluded (*reeser(vaat)-reesel-reesep*), and the two filler items were excluded at random.

Spoken prime	Visual Target	Condition
hambur	HAM	Word Onset
hambuj		Minimal Mismatch
hambup		Maximal Mismatch
rangcuun		Control

Table 5-1 *Example materials Experiments 5.5 and 5.6*

The 32 remaining experimental triplets were used to construct quadruplets of prime-target pairs (see also Table 5-1). The embedded words of the triplets, for example *ham* from *hambur-hambuj-hambup*, were used as visual targets in the three experimental conditions, Word Onset, Minimal Mismatch and Maximal Mismatch, and in a control condition. In the control condition the response to the same visual target was measured when it was preceded by a different, unrelated, item from the stimulus set. Half of the control auditory primes were Word Onsets and half of them were Maximal Mismatches.

Filler prime-target pairs were constructed in a similar way. To avoid a response bias to word onsets, filler pairs were included with existing Dutch words as targets. Twelve matched filler triplets were paired with twelve visual nonword targets, of which six were phonologically related to the second syllable of the auditory prime, and the other six were unrelated. Phonologically related visual targets shared all phonemes but one with the second syllable of the auditory prime. They were included to avoid participants strategically responding to an overlap between the first syllable of the auditory stimulus and the visual target (see also Vroomen & de Gelder, 1995; Zwitserlood, 1996). Eight matched filler triplets were combined with existing Dutch target words that were phonologically unrelated. The other twelve matched filler triplets served as auditory stimuli on twelve catch trials. Similarly, the free fillers were paired with 20 nonwords, 10 phonologically related and 10 unrelated, and 8 words, with the 4 remaining free fillers serving as auditory stimuli for the catch trials. A total of sixteen auditory stimuli (12 matched fillers and 4 free fillers) were used for the catch trials. For half of these 16 auditory primes letter strings were created that formed an adequate orthographic description for the first syllable of the auditory prime (e.g. ZUIF for *zuifkaaf*), while for the other 8 auditory primes letter strings were generated that did not form a possible written version of the first syllable of the prime (e.g. KUIM for *ruumfoum*).

The four different visual target-prime pairs for the experimental items and the matched fillers were counterbalanced across four different versions of the experiment, so that each visual target appeared only once. Each member of an experimental quadruplet or matched filler quadruplet and all free fillers were presented in the same pseudo-random order in all four versions. Each version therefore contained a total of 32 critical pairs (8 in each condition), 20 matched filler pairs (5 in each condition) 28 unrelated filler pairs and a total of 16 catch pairs (12 matched fillers and 4 free fillers). The visual targets were phonologically unrelated to the auditory prime in half of the priming trials, 8 in the control condition of the critical pairs, 14 in the matched fillers and 18 in the free fillers. In the other 40 priming pairs the visual target was either embedded in the auditory prime, i.e. 24 in three experimental conditions, or largely overlapping, 16 in two filler conditions (8 each). The sixteen catch trials were equally spread across the list.

Subjects

Forty-eight subjects participated in four groups of 12 subjects for each version. Participants were again recruited from the Max Planck subject pool. They were paid for participation. No participants reported any vision or hearing disorders and none of them had participated in any of the other experiments.

Procedure

Subjects were tested in groups of four or less. The presentation of the stimuli, the timing of the manual responses and the collection of the data was again performed by the NESU experimental system. Participants were each seated in front of a computer screen, in four different compartments of the experimental room that were separated by sound attenuating screens. The auditory stimuli were played binaurally via Sennheiser headphones. The visual stimuli were presented centered on the computer screen in white against a black background. These letter strings were presented unmasked in 36 point Arial font for 1000 ms. The computer screens were also used to indicate the beginning and the end of the experiment and the pause between the practice block and the experimental block. The participants received instructions to listen to a sequence of nonsense bisyllables. While the auditory stimulus was being presented, two things could happen on the computer screen. If a letter string was presented (always at the offset of the embedded word), participants were asked to perform a lexical decision task on the letter string. They were asked to press a green response button marked "JA" (yes) with their preferred hand as quickly as possible, whenever they saw an existing Dutch word on the screen. If the letter string that was presented on the computer screen was not a real Dutch word, they were asked to press a red button marked "NEE" (no) with their other hand. On the catch-trials a warning signal consisting of three asterisks, "***", was presented visually for one second, again starting at the offset of the embedded word. This warning signal was immediately followed by the visual presentation, for one second, of a letter string. In this case, the participants were asked to press the green response button marked "JA" with their preferred hand, as quickly as possible, if the letter string on the screen formed an adequate description of the beginning of the auditory prime. If the letter string that was presented on the computer screen was not a good description, they were asked to press the red button marked "NEE" with their other hand. The 96 trials of the main experiment were preceded by a block of 9 practice trials.

Results and discussion

Data analysis, Experiments 5.5-5.6

Lexical decision latencies were measured from the onset of the visual targets on the computer screen. Responses slower than 1200 ms and cases where subjects did not respond at all were treated as missing data. Missing data points were again replaced by the relevant mean (per subject or per item) per condition.

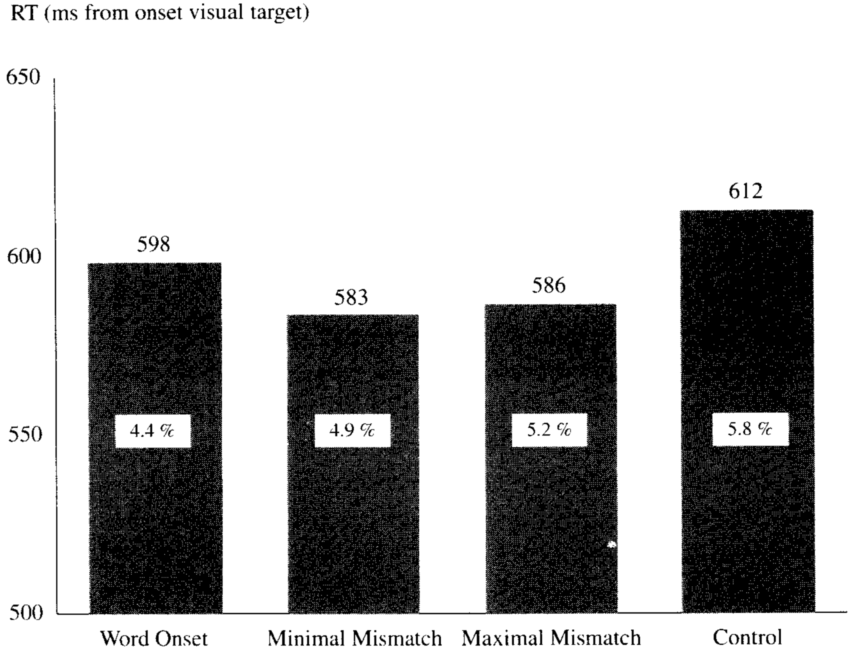

Figure 5-6 *Average lexical decision latencies and error rates (inside the columns) for Experiment 5.5. Results are given for each condition: Word Onset (hambur(ger)-HAM), Minimal Mismatch (hambuj-HAM), Maximal Mismatch (hambup-HAM), and Control (rangcuun-HAM). Visual targets were presented at the offset of the embedded word.*

Results, Experiment 5.5

Analyses of the performance on the secondary task showed that all subjects responded correctly on more than 80% of the catch trials. This suggests that they were indeed attending to the auditory stimuli.

Three items where subjects failed to make the correct lexical decision to the target word in more than 25% of the cases were excluded from the analyses, leaving 29 experimental triplets. The differences between all four conditions, Control, Word Onset, Minimal Mismatch and Maximal Mismatch, were very small (see Figure 5-6 for the mean results from the subjects analysis). The subject analysis of the RTs suggested that there was an overall effect of prime type, $F1$ (3,132) = 2.76, $p < .05$, $MSE = 2904$. This effect, however, was not significant by items, $F2(3,84) = 1.85$, $p > .1$. Post hoc Tukey HSD tests ($p < .05$) by subjects showed that only the Minimal Mismatch was significantly different from the Control baseline. However this difference did not reach significance in the posthoc analyses by items (all p's > .05). Also in the error data no reliable differences were found between the unrelated control and the three related conditions ($F1$ and $F2$, $p > .1$).

In summary, no reliable priming effects were found when the visual target was presented at the offset of the embedded word. A possible explanation of the absence of priming could be given in terms of the time-course of the competition process. If visual processing starts at the offset of the embedded word in the auditory stimulus, the embedded word (e.g. *ham*) might not yet be sufficiently activated in any of the three critical conditions and, therefore might not prime visual lexical decision. Experiment 5.5 could be tapping into the process of competition too early (visual processing starting somewhere before "P" on Figure 5-1) such that only very weak or no priming is found. To investigate whether priming and, possibly, differential competition effects arise at a later stage of processing, a second priming experiment was conducted.

EXPERIMENT 5.6

In Experiment 5.6 the visual stimuli, both for the experimental trials and the catch-trials were presented at a later point relative to the auditory prime: at the offset of the experimental stimulus.

Method

Subjects

Forty-eight subjects from the Max Planck subject pool took part. Twelve subjects were assigned to each of the four versions. Subjects that already participated in any of the previous experiments were excluded.

Materials and Procedure

The materials were identical to those that were used in Experiment 5.5. The procedure was identical to that used in Experiment 5.5 except that the visual stimuli, both the letter strings on the identity priming trials and the warning signal on the catch trials, were now presented at the offset of the second syllable.

Results and discussion

Subjects again responded correctly on more than 80% of the catch trials. Two items where subjects failed to make the correct lexical decision to the target word in more than 25% of the cases were excluded from the analyses, leaving 30 experimental triplets.

The differences between all four conditions, Control, Word Onset, Minimal Mismatch and Maximal Mismatch, were much larger than in Experiment 5.5 (see Figure 5.7 for the mean results from the subjects analysis). The ANOVAs of the reaction times yielded a highly significant overall effect of the relation between prime and target ($F1\ (3,132) = 26.83$, $p < .01$, $MSE = 2260$; $F2(3,87) = 4.58$, $p < .01$, $MSE = 2664$). The results of post hoc Tukey HSD tests ($p < .05$) revealed an intermediate effect for the Minimal Mismatch condition. By subjects, the Maximal Mismatch condition was reliably different from the Control condition and also the Minimal Mismatch condition differed significantly from the Control condition. By items, the only significant difference was between the Maximal Mismatch condition and the Control condition. The Word Onset condition did not show reliable priming compared to the baseline control (Item means: Word Onset = 564 ms; Minimal Mismatch = 559 ms; Maximal Mismatch = 542 ms; Control = 591 ms). No reliable differences were found in similar analyses of the error rates.

For closer inspection of the facilitation by the three different related auditory primes, difference scores were computed by subjects and by items. A difference score by subjects was defined as the mean RT of a specific participant to the target words preceded by related primes minus the mean RT to the target words preceded by unrelated primes (the baseline control condition) for all three conditions separately. Likewise, the difference score by items was defined as the mean RT to a specific target word preceded by a related prime minus the mean RT to the same target word preceded by an unrelated prime for all three related conditions separately. The difference scores were again submitted to ANOVAs both by subjects and by items. The ANOVAs showed no reliable overall effect of the three related priming conditions ($F1$ (2,88) = 2.69, .05 < p < .1, MSE = 2409; $F2$(2,58)= 2.22, p > .1, MSE = 1759). This absence of reliable differences of the differences suggested that all related primes facilitated lexical decision, even in the Word Onset condition.

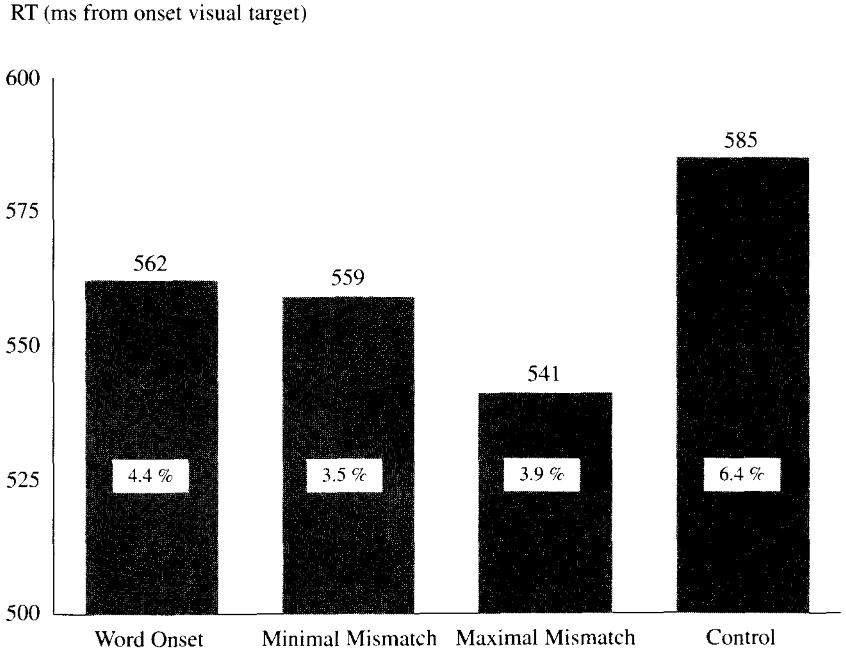

Figure 5-7 *Average lexical decision latencies and error rates (inside the columns) for Experiment 5.6. Results are given for each condition: Word Onset (hambur(ger)-HAM), Minimal Mismatch (hambuj-HAM), Maximal Mismatch (hambup-HAM), and Control (rangcuun-HAM). Visual targets were presented at the offset of the bisyllabic stimulus.*

The difference scores per priming condition were also tested against zero using *t*-tests. These analyses indicated that the visual lexical decision to words was most facilitated by preceding related auditory primes in the Maximal Mismatch condition ($t1(47) = -4.5$, $p < .001$; $t2(29) = -3.47$, $p < .01$) and not reliably facilitated in the Word Onset condition ($t1(47) = -2.23$, $p < .05$; $t2\ (29) = -1.68$, $p > .1$). The Minimal Mismatch condition again appeared to be an intermediate case ($t1(47) = -2.58$, $p < .05$; $t2(29) = -1.97$, $.05 < p < .01$).

In summary, differential priming effects were found for the Word Onset, the Minimal Mismatch and the Maximal Mismatch conditions when the visual target was presented at the offset of the embedded word. The lexical decision on a visual target, for instance HAM, was not reliably facilitated by related auditory primes that formed the beginning of a longer word like *hambur.* This absence of priming in the Word Onset condition can be explained as a competition effect. The longer word (*hamburger*) competes with the embedded word (ham), suppressing the activation of the embedded word, thus blocking the identity priming effect. In the Maximal Mismatch condition, the input [hɑmbʉp] mismatches with *hamburger* to block activation of that word, so there is little or no competition, and lexical decisions to HAM are relatively easy. Furthermore, the present results suggest that the Minimal Mismatch condition is an intermediate case. The input [hɑmbʉj] mismatches with *hamburger* to a smaller degree than [hɑmbʉp] and, consequently, blocks the activation of that word to a smaller degree. In other words, there is still some competition from the longer word candidate but not enough to completely cancel out the priming effect. This suggests that late occurring single feature mismatches (e.g. changes in place of articulation) are tolerated more than those mismatching by two features (e.g. changes in place and manner).

GENERAL DISCUSSION

The main aim of this research was to investigate the influence of late occurring mismatching information on competition between lexical candidates. Overall, the results suggest that listeners are relatively intolerant to late mismatches at a phonetic level. Furthermore, the word recognition system seems to be just as sensitive to late occurring mismatching information as to word-initial mismatching information.

Results from word-spotting (Experiment 5.4) and from identity priming (Experiment 5.6) demonstrated that listeners deal with phonetic mismatches in a way that is independent of the location of the deviant information. In Experiment 5.4 participants were asked to listen to a list of bisyllabic nonsense items and to try to spot words embedded both at the beginnings and at the ends of the stimuli. The results from Experiment 5.4 not only provide converging evidence for the relative sensitivity of the spoken word recognition system to word-initial phonetic mismatches (Chapter 4), but also extend these findings to later occurring mismatches. As in the previous study, listeners found it more difficult to spot words, for example *les*, that were embedded in a longer utterance that formed the beginning of an existing Dutch word, for example *choles* excised from *cholesterol*, than in words that did not form the beginning of a word in Dutch (e.g. *les* in *foles* excised from *folesterol* and *les* in *boles* excised from *bolesterol*). As before, the data also indicate that mismatches by only one feature (i.e. place of articulation) are tolerated more than mismatches by several features (i.e. place, voice and manner). Hearing *foles* from *folesterol* activates *cholesterol* to a smaller extent than the word onset *choles* from *cholesterol*, but to a larger extent than *boles* from *bolesterol*.

Similar results were obtained for late mismatches. Listeners found it more difficult to spot words, for example *ham*, that were embedded in a longer utterance that formed the beginning of an existing Dutch word, for example *hambur* excised from *hamburger*, than in words that did not form the beginning of a word in Dutch (e.g. *ham* in *hambuj* excised from *hambujger* and *ham* in *hambup* excised from *hambupger*). Furthermore, the data indicate that late mismatches by only a single feature (i.e. place of articulation) are tolerated more than mismatches by two features (i.e. place and manner). Hearing *hambuj* from *hambujger* activates *hamburger* to a smaller extent than the word onset *hambur* from *hamburger*, but to a larger extent than *hambup* from *hamburger*. This mismatch effect did not interact with the position of the mismatch.

This graded effect of late occurring mismatches was confirmed by the results from Experiment 5.6. In Experiment 5.6 the influence of late mismatches was investigated in a cross-modal identity priming task. It was demonstrated that listeners found it easier to make a lexical decision to words, for example HAM, that were presented visually right after hearing a fragment with that word embedded in it (e.g. HAM after hearing *hambup*) than after hearing an unrelated fragment (e.g. HAM after hearing *rangcuun*). No reliable facilitation was found when the related fragment also formed the beginning of an existing word in Dutch (e.g. HAM after hearing *hambur*). This absence of priming was interpreted as an effect of competition from the longer word candidate. Furthermore, the data indicate that mismatches by only one feature (i.e. place of articulation) are tolerated more than mismatches by two features (i.e. place and manner). Hearing *hambuj* activates *hamburger* to a smaller extent than the word onset *hambur* but to a larger extent than *hambup,* causing an intermediate priming effect.

Taken together, the results from Experiment 5.4 and Experiment 5.6 support the hypothesis that early and late mismatching information do not have a different status for the word recognition system. In Experiment 5.4, the competition effects that were found earlier for word-initial mismatches (Chapter 4) are observed for both early and late mismatches. In this previous study, it was demonstrated that the word-spotting paradigm was tapping into the right stage of the recognition process to observe competition effects from mismatching candidates even when participants were asked to spot words in a single target location. However, in the first three word-spotting experiments of the present study this was no longer the case.

In Experiment 5.1 subjects found it as hard to spot words that were embedded at the beginning of an existing Dutch word (e.g. *ham* in *hambur*) as to detect words that were embedded at the beginnings of bisyllables that did not form the onset of a real Dutch word (e.g. *ham* in *hambuj* or *ham* in *hambup*). The absence of a competition effect in Experiment 5.1 was explained in terms of the time-course of the competition process. Changing the position of the mismatching information in the input and also changing the position of the embedded word in the embedding fragment might have changed the time-course of the recognition process in such a way that competition effects were no longer observed when participants were only listening for finally embedded words. Simulations using the most recent version of the SHORTLIST model (Norris, 1994) that is adapted to include the Possible Word Constraint (Norris et al., 1997) were run to evaluate whether competition effects were to be expected for late occurring mismatching information and, is so, how they evolve in time. These simulations predicted that it should be possible to obtain competition effects of late mismatches, provided that the experimental task is tapping into the right stage of processing. In line with McQueen et al. (1994), it was argued that specification of the position of the embedded word allowed the listeners to respond too fast for competition effects to have emerged. Therefore, in Experiments 5.2 and 5.3, a different recording method was used.

In Experiments 5.2 and 5.3 the complete longer word candidate (e.g. *hamburger*) and its mismatching derivatives (e.g. *hambujger* and *hambupger*) were recorded and bisyllabic stimuli were generated from these items. On the basis of previous results (Chapter 4) it was hypothesized that applying this procedure could result in the emergence of a differential competition effect at a later point in time. The data from Experiment 5.2 showed a mixed pattern of results. A large proportion of listeners found it easier to spot words, for example *ham*, that were embedded in a longer utterance that formed the beginning of an existing Dutch word, for example *hambur*, than in words that did not form the beginning of a word in Dutch (e.g. *ham* in *hambup*). This effect in the direction opposite to that predicted was attributed to a bias in the stimulus set. The participants could have been strategically responding to word onsets, since all of the targets and all of the matched fillers were word onsets up to the final consonants.

In Experiment 5.3 fillers were added to change the proportion of word onsets. It was predicted that adding these fillers that did not form the beginning of any word would make the 'word onset'-bias in the stimulus set less apparent. However, the results from Experiment 5.3 indicate that changing the proportion of word onsets did not have the intended effect. Although the effects found in Experiment 5.2 were attenuated in Experiment 5.3, they did not disappear entirely: error rates in the word onset condition remained lower than in the other two conditions. To sum up, in Experiments 5.2 and 5.3 (and also perhaps to some extent in Experiment 5.1) possible effects of competition were obscured by the strategic behaviour of the participants. It was suggested that when subjects were only listening for word-initial targets, the embedded words were recognized via the recognition of the embedding words. The current results show that word-spotting is, in certain conditions, susceptible to strategic processes. For instance, orthography can be used to accomplish the task in languages with a good correspondence between letters and sounds (cf. Dijkstra et al., 1995). This not only underlines the need for a careful selection of materials (Cutler, 1981; Forster, submitted), but the results from Experiment 5.4 also suggest that using different target positions in a single experiment may help to avoid strategic behaviour. Furthermore, it has been demonstrated that using a different recording method may change the lexical extent of the bisyllabic stimuli in such a way that a bias in the materials becomes more apparent to the listener. It was argued in Chapter 4 that if the first two syllables are taken from a token of the complete embedding word, there is more information in support of the longer candidate word, relative to the support for the embedded word when the two syllables are produced in isolation. This might explain why the strongest evidence of the strategy was observed in Experiments 5.2 and 5.3. In general, with effects found in laboratory tasks it is always important to control for possible strategies. Every paradigm has its own merits and limitations.

In previous research (Vroomen & de Gelder, 1995) and in the present study, identity priming has been shown to be a paradigm that can be used in addition to word-spotting for the investigation of competition effects. However, in Experiment 5.5 no effects of competition were found. Listeners found it equally difficult to make a visual lexical decision to words, for example HAM, that were preceded by a related fragment prime, for example *hambur, hambuj* or *hambup*, than to the same word that was preceded by an unrelated fragment prime, for instance *rangcuun*. In other words, none of the related embedding fragments primed lexical decision when the embedded word was presented visually at the offset of the embedded word in the auditory stimulus. This suggests that at this point in time the embedded word was not sufficiently activated to prime lexical decision in any of the three conditions. Shifting the Stimulus Onset Asynchrony (SOA) in Experiment 5.6 produced the predicted priming effects. The appropriate timing of the visual stimulus relative to the auditory stimulus (which is manipulated to track the time-course of recognition) is crucial for effects of competition in identity priming to become observable (see also Chapter 4).

The word-spotting results for final targets in Experiment 5.4 provide converging evidence for the relative sensitivity of the spoken word recognition system to word-initial phonetic mismatches. In line with earlier research (Bölte, 1997; Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997; Marslen-Wilson, 1993; Marslen-Wilson, Moss, & van Halen, 1996; Marslen-Wilson, & Zwitserlood, 1989; Chapter 4) it was shown in Experiment 5.4 that an input that mismatches by a single feature is not treated as an instance of the original word onset fragment. Furthermore it was shown that single feature mismatches, only differing in place of articulation, were tolerated more than mismatches by several features (place, manner and voice). This is consistent with earlier findings that demonstrated graded effects of word-initial mismatch (Connine, Blasko, & Titone, 1993; Connine, Titone, Deelman, & Blasko, 1997; Chapter 4).

In addition, the present results extend the previous findings for word-initial mismatches to mismatching information at later positions. In line with earlier research (Marslen-Wilson, 1993) it was shown in Experiments 5.4 and 5.6 that an input that mismatches at offset by a single feature is also not treated as an instance of the original word onset fragment. Furthermore it was shown that, like early mismatches, late single feature mismatches, only differing in place of articulation, were tolerated more than mismatches by two features (place and manner). This is consistent with earlier findings that demonstrated similar effects of mismatching information in word-medial position (Connine, 1994; Connine et al., 1993).

In order to account for the data in the present study, models of spoken word recognition would need to incorporate some kind of mechanism that performs flexibly with respect to the amount of mismatching information. In other words, single feature mismatches should be treated differently from several feature mismatches.

Both in the COHORT model (Marslen-Wilson, 1987, 1990; Marslen-Wilson & Welsh, 1978) and in the SHORTLIST model (Norris, 1994; Norris et al., 1997), mismatching information is used explicitly for the selection of the appropriate candidate word. In the COHORT model any one-feature or larger mismatch prevents possible words from being activated as candidates altogether. The current study shows that single feature mismatch should not block lexical access entirely, as was argued for by Marslen-Wilson et al. (1996). These results therefore challenge the Cohort model. Multi-feature mismatch, on the other hand, at least under the carefully controlled circumstances of a laboratory experiment (high quality speech presented in sound-attenuated booths over high-quality headphones) does seem to be enough to prevent words attaining levels of activation sufficient to influence behavioural measures.

In SHORTLIST (Norris, 1994; Norris et al., 1997), mismatching segments in the input lower the bottom-up support for lexical candidates, making it unlikely that mismatching candidates will enter the 'shortlist' and thus that they will probably not actively compete for recognition. In the current version of the model graded effects of mismatch can be modelled by using mid-class phoneme transcriptions for inputs that should not act to block lexical access completely. As was shown in the current study, this provides a useful tool to evaluate effects of mismatching information at different stages of processing. However, the current version of the SHORTLIST model does not incorporate a psychologically plausible mechanism to deal with mismatching information in a flexible fashion.

In the TRACE model (Elman & McClelland, 1988; McClelland 1979, 1987, 1991; McClelland & Elman, 1986; McClelland & Rumelhart, 1981) mismatching information does not have a direct effect on lexical access. It might have been modelled by inhibitory phoneme-to-word connections (as was also discussed by McClelland and Elman, 1986, p. 55-57), but they decided against it, arguing that word-to-word inhibition would have the same effect.

However, this is only true if the mismatching input provides information in favour of another word, rather than a nonword. If the mismatching information supports a nonword, then in cases where the input shares no features with the relevant phoneme of the possible word the mismatching information will be treated similarly to the absence of matching information by the model. If the mismatching input shares one or more features (note that single feature mismatches are in fact matches on several other features), mismatching inputs will not block lexical access entirely. In this respect, the TRACE model seems to perform better than the other two models. However, the evidence in favour of a direct use of mismatching information from the numerous studies demonstrating that the word recognition system is rather intolerant of mismatching information seems to make a TRACE-type model less plausible.

In conclusion, it has been argued that spoken word recognition is equally sensitive to late phonetic mismatches as to word-initial phonetic mismatches. Only small amounts of mismatching information are tolerated, independent of the location of the mismatching information. The observed mismatch effects in word-spotting and identity-priming suggest a graded effect of the amount of mismatch. Mismatch is tolerated more when a segment of an utterance mismatches a candidate word by a single feature than when it differs by two features. The drunk will probably get his nicotine fix by asking for a *cigalette*, but will probably remain deprived if he asks for a *cigapette*.

SUMMARY AND CONCLUSIONS

CHAPTER 6

Before speech can be recognised as a sequence of words that conveys a meaningful message from speaker to listener, two important problems have to be solved by the listener: the problem of segmentation and the problem of variation. In the present study 18 experiments were reported. Eight experiments (Chapters 2 and 3) were concerned with the segmentation problem; they examined the use of sequential probabilities in segmentation. Ten experiments dealt with the problem of variation through the examination of the role of mismatching information in spoken word recognition (Chapters 4 and 5).

SEQUENTIAL PROBABILITIES IN SEGMENTATION

The speech signal is continuous. Word boundaries are not reliably marked in spoken language. However, in order to understand what is said the listener has to segment the utterances into words. How do people solve this problem of segmentation?

In the first experiment (Chapter 2), a word-spotting task was used to investigate whether listeners use the relative probability of occurrence of consonant clusters as word onsets or offsets in segmentation. The six consonant clusters that were tested can all appear both at the beginning and at the end of a word in Dutch. However, they differed in the relative likelihood of occurring in a particular position. For example, the cluster [sp] is onset dominant. That is, it occurs more often at the beginning of a word than at the end of a word. In contrast, the cluster [ks] occurs more often at the end of a word than at the beginning of a word. The results provided only weak evidence in favour of the hypothesis that listeners use this kind of information for parsing the speech stream into words. This failure to obtain strong effects of relative positional cluster probabilities was held to be due to two factors. First, the absolute probabilities of occurrence of these consonant clusters might have obscured the predicted effects of relative probability. For example, although the consonant cluster [sp] occurs much more often at the beginnings of words than at the ends, it is still the final sound sequence of no fewer than 52 words in Dutch.

Second, the special status of the segment [s] in complex consonant clusters might also have interfered with the hypothesized effects of sequential probability. Therefore, in a series of seven follow-up experiments absolute frequencies of occurrence were investigated, using single-consonant onsets and codas.

In Chapter 3, the results are reported of five experiments using the word-spotting task to test whether the frequency of occurrence of word-initial CV or word-final VC strings contributes to the task of extracting Dutch words from larger stretches of speech. These CV and VC strings were placed either at the edge of the stretch abutting the word (e.g. *boom-**dif***) or at the edge of the word abutting the remnant of the speech stream (e.g. ***heup**-juik*). The hypothesis tested in these experiments was that listeners would be better at spotting words if the sequential probabilities of these strings were higher than if the probabilities were low.Two control experiments were run to determine whether sequential probability differences have an effect on lexical decision. The purpose of these latter experiments was to make sure that any sequential probability effects in the word-spotting task reflected the use of sequential probabilities in segmentation itself and not in deciding whether a stretch of segments is a word or not.

The results from Chapter 3 demonstrated that listeners indeed use knowledge of sequential probabilities in segmentation. However, these findings also suggested two ways in which listeners are constrained in their use of this kind of distributional information. The probability of a sound sequence influenced segmentation more when the sequence occurred within the target words. Furthermore, position-specific sequential probability information seems to be exploited most effectively when it cues the onsets of words.

As was discussed in the introduction to Chapter 3, other work has shown that differences in sequential probability have profound effects in other speech and word perception tasks (Aslin, Saffran & Newport, 1998; Jusczyk, Luce & Charles-Luce, 1994; Mattys, Jusczyk, Luce, & Morgan, 1999; McQueen & Pitt, 1996; Pitt & McQueen, 1998; Pitt & Samuel, 1995; Saffran, Aslin & Newport, 1996; Vitevitch & Luce, 1998; Vitevitch, Luce, Charles-Luce & Kemmerer, 1997) . These findings have led the developers of infant and adult models of spoken word recognition to argue that knowledge of these probabilities should be included in their models. Together, the findings in Chapter 2 and 3, however, provide evidence of limitations on the use of this knowledge by human listeners. That there were limitations in the use of this knowledge is not surprising, since the cue that it provides to word boundaries is relatively weak.

A low probability onset is still an onset of a number of words; a high probability onset is likely to be but is not necessarily the onset of a word. Furthermore, other cues were available too, phonotactics in particular.

Interaction of boundary cues

Given that there are a multitude of segmentation cues, an important issue for future research is to establish how and when they interact. In the present study, the use of sequential probability information as a cue towards the location of word boundaries was investigated in conjunction with only one other cue to segmentation, namely that provided by phonotactics. Previous studies have shown that the integration of boundary markers can vary enormously depending on the situation. For example, opposite findings are reported for adult and infant listeners with respect to the interaction of prosodic and phonotactic information. When both types of cues are present in the speech stream, infants rely more strongly on the metrical information (Mattys et al., 1999), whereas adults seem to rely more strongly on phonotactics (McQueen, 1998). Sequential probability information might become more important in the absence of other cues to word boundaries, and it may interact differently with other cues, such as those provided by prosody, than it does with phonotactic cues. As a first approach to this issue, one might consider running meta-analyses on the huge amount of experimental data that is already available on the use of particular boundary cues in isolation (cf. Pitt & Samuel, 1993). Experiments need to be designed to investigate directly whether and how different boundary cues interact in the process of segmenting speech and what their relative importance is.

Acquisition and storage of sequential probabilities

Current implementations of models of adult spoken word recognition lack a mechanism to account for the findings in Chapter 2 and 3. Such a mechanism would incorporate some kind of prelexical analysis of the speech signal that would result in the accumulation and storage of distributional information. Several mechanisms have been proposed in the literature, but none has been fully implemented.

In the original SHORTLIST paper, Norris (1994) proposed a theoretical solution in the form of a Jordan-type connectionist network that can learn and store sequential probability information (see also Norris, 1990 and Norris, 1993). This type of recurrent network can learn and store sequential dependencies in the input through the operation of context units that provide the network with a memory of the previous state of the network and hidden units that integrate the current input with the previous state of the network (see also Elman, 1990). However, for the sake of speed and simplicity in the SHORTLIST simulations, this recurrent front-end to the process of lexical competition was never implemented and its output was instead simulated by an exhaustive search through a machine-readable dictionary.

As was mentioned in the discussion of Chapter 3 a different approach is taken by Vitevitch and Luce (1999). They propose that an adaptive resonance framework (Grossberg, Boardman, & Cohen, 1997) can account for both the lexical and sublexical levels of speech processing. Within this framework, sequential probability information is stored in the form of sublexical chunks in working memory which evolve through repeated exposure.

An alternative solution for learning and storing distributional information might be provided by attractor-type connectionist models with distributed representations and modifiable weights, for example, the Hinton, Plaut and Shallice models (Hinton & Shallice, 1991; Plaut & Shallice, 1993) which were proposed for impaired reading. Learning sequential probabilities in an attractor-type model can be seen as the development of global sublexical attractor basins which evolve in the multidimensional space of all possible activation patterns through encounters with similar items. These global sublexical attractor basins may help the recognition system to settle quickly into the appropiate region of more local lexical attractor basins (cf. Prager, Harrison & Fallside, 1986).

Independent of which approach is taken, the challenge for future models of adult spoken word recognition is that they should operate on a more realistic input. Most models currently operate on an idealised representation of the speech input. These idealised representations do not reflect the distributional information that has been shown to be exploited by the human listener in the segmentation of spoken language.

Acquisition of a lexicon

Most models for adult spoken word recognition oversimplify the problem of segmentation by assuming that segmentation is a by-product of word recognition. In these models, locating the word boundaries crucially depends on knowing the word. This may be a satisfactory mechanism for adult listeners, but it cannot be the correct mechanism by which infants first learn to segment speech. In the introduction to Chapter 3, a number of studies on speech segmentation in infants were reported. These studies clearly show that segmentation does not crucially depend on lexical knowledge. Brent and Cartwright (1996) proposed a computational model that performs segmentation of an input on the basis of distributional information, in the absence of a lexicon of word candidates. Because the model also operates through the application of the principle of possible word parsing, it is closely related to the SHORTLIST model for adult spoken word recognition. The solution Brent & Cartwright propose is purely distributional, while Jusczyk and collaborators argue that distributions are just one of a multitude of cues (Jusczyk, Cutler, & Redanz, 1993; Jusczyk, Friederici, Wessels, Svenkerud, & Jusczyk, 1993; Jusczyk, Luce & Charles-Luce, 1994; Mattys, Jusczyk, Luce, & Morgan 1999; Turk, Jusczyk, & Gerken, 1995). The idea that segmentation occurs through the integration of multiple boundary cues is also closely related to the SHORTLIST idea. In the SHORTLIST model, several boundary cues (e.g. metrical, phonotactic and silence boundary markers can be specified in the input) are integrated through the operation of the Possible Word Constraint. Ideally, infant and adult approaches should be coupled more closely. The SHORTLIST model seems to provide a nice framework for this unification.

Mismatch and Competition

There is huge variation in the realisation of speech sounds. Nevertheless, human speech perception is remarkably stable. How do listeners deal with this problem of variation?

In order to adjust flexibly to changing context one would expect some degree of flexibility in the process of word recognition with respect to imperfect realisations of words. In Chapters 4 and 5, six word-spotting experiments and four cross-modal identity priming experiments were used to investigate how tolerant spoken word recognition is to mismatching information, that is, to the situation where phonological information in the speech signal mismatches with phonological information stored in the mental lexicon. Single feature phonetic mismatches were contrasted with mismatches on several phonetic features. Sensitivity was measured indirectly through the evaluation of competition between a word that might be sufficiently activated in spite of mismatching information and other lexical candidates. In four experiments (Chapter 4), the mismatching information was located at the beginning of a lexical candidate. In the other six experiments (Chapter 5), the mismatching information was presented at a later position in the word.

Taken together, the results suggest that listeners are rather intolerant of mismatches. This is not surprising if we consider the extreme consequence of this type of flexibility: we would hear things that were not spoken and hallucinate instead of understanding speech. However, the findings also demonstrated that single feature mismatches are less distinctive than mismatches involving several features, irrespective of the position of the mismatching information.

As was discussed in the introduction to Chapter 5, other work has shown that the tasks used in other experimental investigations of the role of mismatching information are susceptible to response strategies. In the first three word-spotting experiments in Chapter 5, it was demonstrated that, in certain conditions, listeners in an auditory task like word-spotting can be biased towards response strategies that reflect orthographic rather than auditory processing.

There is therefore a danger that word-spotting materials can invoke strategic processing. Several other factors were mentioned in the discussions of Chapter 4 and Chapter 5 which can influence the likelihood of obtaining effects of mismatching information in an experimental setting. Mismatch effects express themselves through the activation of lexical candidates. In line with Connine (1994), it was argued that the observation of mismatch effects might depend on the *lexical extent* of the stimuli, that is, the amount of information in support of a word candidate. In all studies, lexical extent is modulated by a second important factor: the listening conditions. It was argued that if you are talking to a person in a noisy bar, you are more likely to be more tolerant towards the information that reaches the ear, than when you are seated in a sound attenuated booth.

Together, the findings in Chapter 4 and 5 suggest limitations on the flexibility of human listeners in the processing of speech. Furthermore, these findings motivate a theoretical explanation for the absence versus presence of mismatch effects in experiments on speech comprehension. At least under laboratory listening conditions, it appears that listeners tolerate single feature mismatches in lexical access, but no more than that.

Models of spoken word recognition

Like speech, spoken word recognition unfolds in time. The temporal nature of the recognition process poses an additional problem to the psycholinguist: when does an effect occur? In the present study, and elsewhere, models of spoken word recognition have been shown to provide an excellent tool to evaluate possible effects at different points in time. In order to account for the data in the present study, models of spoken word recognition would however need to incorporate some kind of mechanism that performs flexibly with respect to the amount of mismatching information. In other words, single feature mismatches should be treated differently from mismatches on several features. Furthermore, the system should be able to adjust flexibly to changes in listening conditions and lexical extent.

The findings in Chapters 4 and 5 showed that single feature mismatch should not block lexical access entirely, as was argued by Marslen-Wilson and collaborators (e.g., Marslen-Wilson, Moss & van Halen, 1996). These results therefore challenge the COHORT model (Marslen-Wilson, 1987, 1990; Marslen-Wilson & Welsh, 1978). In the COHORT model any one-feature or larger mismatch prevents possible words from being activated as candidates altogether. Multi-feature mismatch, on the other hand, at least under the carefully controlled circumstances of a laboratory experiment, does seem to be enough to prevent words attaining levels of activation sufficient to influence behavioural measures.

The SHORTLIST model has been shown to be able to simulate the current empirical data through the operation of predefined mid-class phoneme transcriptions in the input. However, this is not a psychologically plausible mechanism. As I will argue below, a more plausible account of the input representation in SHORTLIST is required for the model to be able to deal with mismatching information.

In the TRACE model, mismatch is not used directly in the recognition process. The TRACE model was shown to perform better than the other two models in the modelling of graded effects of mismatch, through the operation of a featural input level. However, the evidence in favour of a direct use of mismatching information, that is, that a multiple-feature mismatch on one phoneme is sufficient to block lexical access entirely, makes a TRACE-type model less plausible.

The same conclusion as was drawn from the research on segmentation can also be drawn here: *the* challenge for future models of adult spoken word recognition is that they should operate on a more realistic, perhaps probabilistic input. That is, the input should be one that reflects the results of the sublexical processing which deals with variation. The idealised inputs for current models do not reflect the tolerance of human word recognition to single feature mismatches in imperfect realisations, nor do these representations in any way reflect how listening conditions could modulate the listener's flexibility.

THE EXPERIMENTER'S DILEMMA

The psycholinguist is interested in one of the great accomplishments of human cognition: the ability to acquire, comprehend and produce natural language. In order to learn more about this intriguing topic, the researcher has to decompose and narrow down the subject matter (e.g., by focusing on specific topics like segmentation and mismatch in auditory word recognition). Subsequently, the psycholinguist runs carefully controlled experiments with participants from the Nintendo generation (*Shall we play a game?*) and with materials that are selected to avoid any confounds. The results look promising, but what do they tell us about the real thing? In a preliminary field study that was carried out in the bar of the Jonkerbosch Conference Centre in Nijmegen, 25 subjects (served beers for participation) were presented with the following speech stream: "maggikunzigaletjevajuhbietsuh": ("*canibumashigalet*"). The results suggest that even minimal mismatches can seriously damage my health.

REFERENCES

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

Aslin, R. N., Saffran, J. R., & Newport, E. L. (1998). Computation of conditional probability statistics by 8-month-old infants. *Psychological Science, 9,* 321-324.

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.

Brent, M. R., & Cartwright, T. A. (1996). Distributional regularity and phonotactic constraints are useful for segmentation. *Cognition, 61,* 93-125.

Booij, G. (1995). *The phonology of Dutch.* Oxford, Clarendon Press.

Bölte, J. (1997). *The role of mismatching information in spoken word recognition.* Hamburg: Kovac.

Burnage, G. (1990). *CELEX: A guide for users.* Nijmegen, The Netherlands: CELEX.

Cairns, P., Shillcock, R., Chater, N., & Levy, J. (1997). Bootstrapping word boundaries: A bottom-up corpus-based approach to speech segmentation. *Cognitive Psychology, 33,* 11-153.

Chomsky, N. & Halle, M. (1968). *The sound pattern of English.* Cambridge, MA: MIT Press.

Christiansen, M. H., Allen, J., & Seidenberg, M. S. (1998). Learning to segment speech using multiple cues: A connectionist model. *Language and Cognitive Processes, 13,* 221-268.

Cluff, M. & Luce, P. A. (1990). Similarity neighborhoods of spoken two-syllable words: Retroactive effects on multiple activation. *Journal of Experimental Psychology: Human Perception and Performance. 16,* 551-563.

Cole, R. A. (1973). Listening for mispronunciations: a measure of what we hear during speech. *Perception & Psychophysics, 11,* 153-156.

Cole, R. A., & Jakimik, J. (1978). Understanding speech: How words are heard. In G. Underwood (Ed.), *Strategies of information processing* (pp. 67-116). London: Academic Press.

Cole, R. A., & Jakimik, J. (1980). A model of speech perception. In R. A. Cole (Ed.), *Perception and production of fluent speech* (pp. 133-164). Hillsdale, N.J.: Erlbaum.

Cole, R. A., Jakimik, J., & Cooper, W. E. (1978). Perceptibility of phonetic features in fluent speech. *Journal of the Acoustical Society of America, 64,* 44-56.

Connine, C. M. (1994). Vertical and horizontal similarity in spoken word recognition. In C. Clifton, Jr., L. Frazier, & K. Rayner (Eds.), *Perspectives on sentence processing* (pp. 107-120), Hillsdale, N.J.: Erlbaum.

Connine, C. M., Blasko, D. G., & Titone, D. A. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language, 32*, 193-210.

Connine, C. M., Blasko, D. G., & Wang, J. (1994). Vertical similarity in spoken word recognition: Multiple lexical activation, individual differences, and the role of sentence context. *Perception & Psychophysics, 56*, 624-636.

Connine, C. M., Titone, D. A., Deelman, T., & Blasko, D. G. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language, 37*, 463-480.

Cutler, A. (1981). Making up materials is a confounded nuisance, or: Will we be able to run any psycholinguistic experiments at all in 1990? *Cognition, 10*, 65-70.

Cutler, A., Mehler, J., Norris, D., & Segui, J. (1986). The syllable's differing role in the segmentation of French and English. *Journal of Memory and Language, 25*, 385-400.

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.

Cutler, A., Treiman, R., & Ooijen, B. van (1998). Orthografik inkoncistensy ephekts in foneme detektion? In R. H. Mannell & J. Robert-Ribes (Eds.), *Proceedings of the 5th International Conference on Spoken Language Processing* (pp. 2783-2786). Sydney: Australian Speech Science and Technology Association.

Dijkstra, T., Roelofs, A., & Fieuws, S. (1995). Orthographic effects on phoneme monitoring. *Canadian Journal of Experimental Psychology, 49*, 264-271.

Dumay, N., Banel, M. H., Frauenfelder, U. H., & Content, A. (1998). Le rôle de la syllabe: segmentation lexicale ou classification? *Actes des 2èmes Journées d'Etude sur la Parole* (pp. 33-36). Martigny, Switzerland.

Dumay, N., Content, A., & Frauenfelder, U. H. (1999). Acoustic-phonetic cues to word boundary location: evidence from word spotting. *14th International Congress of Phonetic Sciences*. Berkeley, CA: University of California.

Dupoux, E. & Mehler, J. (1992). Unifying awareness and on-line studies of speech: A tentative framework. In J. Alegria, D. Holender, J. Morais, & M. Radeau (Eds.), *Analytic approaches to human cognition* (pp. 59-75).

Amsterdam: Elsevier Science Publishers.

Eimas, P. D. & Corbit, J. D. (1973). Selective adaptation of linguistic feature detectors. *Cognitive Psychology, 4*, 99-109.

Elman, J.L. (1990). Finding structure in time. *Cognitive science, 14*, 179-212.

Elman, J. L. & McClelland, J. L. (1988). Cognitive penetration of the mechanisms of perception: Compensation for coarticulation of lexically restored phonemes. *Journal of Memory and Language, 27*, 143-165.

Forster, K. I. (submitted). How should items be selected in word recognition experiments?

Fowler, C. A. (1984). Segmentation of coarticulated speech in perception. *Perception & Psychophysics, 36*, 359-368.

Frauenfelder, U. H., Segui, J., Dijkstra, A. (1990). Lexical effects in phonemic processing: Facilitory or inhibitory? *Journal of Experimental Psychology: Human Perception and Performance, 16*, 77-91.

Freedman, D. (1992). *Frequency effects in auditory word recognition using the word-spotting task.* Unpublished undergraduate project, Department of Experimental Psychology, University of Cambridge, Cambridge, UK.

Friederici, A. D., & Wessels, J. M. I. (1993). Phonotactic knowledge and its use in infant speech perception. *Perception & Psychophysics, 54*, 287-295.

Goldinger, S. D., Luce P. A., & Pisoni, D. B. (1989). Priming lexical neighbors of spoken words: Effects of competition and inhibition. *Journal of Memory and Language, 28*, 501-518.

Goldinger, S. D., Luce P. A., Pisoni, D. B., & Marcario, J. K. (1992). Form-based priming in spoken-word recognition: the roles of competition and bias. *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1211-1238.

Gow, D. W., & Gordon, P. C. (1995). Lexical and prelexical influences on word segmentation: Evidence from priming. *Journal of Experimental Psychology: Human Perception and Performance, 21*, 344-359.

Grosjean, F. (1980). Spoken word recognition processes and the gating paradigm. *Perception & Psychophysics, 28*, 267-283.

Grossberg, S., Boardman, I., & Cohen, M. (1997). Neural dynamics of variable speech rate categorization. *Journal of Experimental Psychology: Human Perception and Performance, 23*, 483-503.

Hinton, G., & Shallice, T. (1991). Lesioning an attractor network: Investigations of acquired dyslexia. *Psychological Review, 98*, 74-95.

Jusczyk, P. W., Cutler, A., & Redanz, N. (1993). Preference for the predominant stress patterns of English words. *Child Development, 64*, 675-687.

Jusczyk, P. W., Friederici, A. D., Wessels, J., Svenkerud, V. Y., & Jusczyk, A. M. (1993). Infants' sensitivity to the sound patterns of native language words. *Journal of Memory and Language, 32*, 404-420.

Jusczyk, P. W., Luce, P. A., & Charles-Luce, J. (1994). Infants' sensitivity to phonotactic patterns in the native language. *Journal of Memory and Language, 33*, 630-645.

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.

Klatt, D. H. (1989). Review of selected models of speech perception. In W. D. Marslen-Wilson (Ed.), *Lexical representation and process* (pp. 169-226). Cambridge, MA: MIT Press.

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

Lehiste, I. (1960). An acoustic-phonetic study of internal open juncture. *Phonetica, 5 (Suppl. 5)*, 1-54.

Liberman, A. M. & Mattingly, I. G. (1985). The motor theory of speech perception revised. *Cognition, 21*, 1-36.

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

Luce, P. A., & Pisoni, D. B. (1998). Recognizing spoken words: The neighborhood activation model. *Ear & 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. (1980). Speech understanding as a psychological process. In J. C. Siman (Ed.), *Spoken language generation* (pp. 39-67), Dordrecht, The Netherlands: Reidel.

Marslen-Wilson, W. D. (1984). Function and process in spoken word recognition. In H. Bouma & D. G. Bouwhuis (Eds.), *Attention and Performance X: Control of Language Processes* (pp. 125-150). Hillsdale, NJ: Erlbaum.

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

Marslen-Wilson, W. D. (1990). Activation, competition and frequency in lexical access. In: G. T. M. Altmann (Ed.), *Cognitive models of speech* *processing: Psycholinguistic and computational perspectives* (pp. 148-172), Cambridge, MA: MIT Press.

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, N.J.: Erlbaum.

Marslen-Wilson, W. D., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance, 22*, 1376-1392.

Marslen-Wilson, W. D. & Tyler, L. K. (1980). The temporal structure of spoken language understanding. *Cognition, 8*, 1-71.

Marslen-Wilson, W. D. & 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.

Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: the importance of word onsets. *Journal of Experimental Psychology: Human Perception & Performance, 15*, 576-585.

Mattys, S. L., Jusczyk P. W., Luce, P. A., & Morgan, J. L. (1999). Phonotactic and prosodic effects on word segmentation in infants. *Cognitive Psychology, 38*, 465-494.

McClelland, J. L. (1979). On the time relations of mental processes: An examination of systems of processes in cascade. *Psychological Review, 86*, 287-330.

McClelland, J. L. (1987). The case for interactionism in language processing. In M. Coltheart (Ed.), *Attention and Performance XII: The Psychology of Reading* (pp. 3-35). Hillsdale: Erlbaum.

McClelland, J. L. (1991). Stochastic interactive processes and the effect of context on perception. *Cognitive Psychology, 23*, 1-44

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

McClelland, J. L. & Rumelhart, D. E. (1981). An interactive activation model of contexts effects in letter perception: Part 1. An account of basic findings, *Psychological Review, 88*, 375-407.

McQueen, J. M. (1991). The influence of the lexicon on phonetic categorization: Stimulus quality in word-final ambiguity. *Journal of*

Experimental Psychology: Human Perception and Performance, 17, 433-443.

McQueen, J. M. (1996). Word spotting. *Language and Cognitive Processes, 11*, 695-699.

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

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. (in press). Lexical influence in phonetic decision-making: Evidence for subcategorical mismatches. *Journal of Experimental Psychology: Human Perception and Performance.*

McQueen, J. M., & Pitt, M. A. (1996) Transitional Probabilities and phoneme monitoring. In H. T. Bunnell & W. Idsardi (Eds.), *Proceedings of the Fourth International Conference on Spoken Language Processing Vol. 4* (pp. 2502-2505). Wilmington, DE: University of Delaware and Alfred I. duPont Institute.

Mehler, J., Dommergues, J.-Y., Frauenfelder, U. H., & Segui, J. (1981). The syllable's role in speech segmentation. *Journal of Verbal Learning and Verbal Behavior, 20*, 298-305.

Miller, J. L. & Liberman, A. M. (1979). Some effects of later-occurring information on the perception of stop consonant and semivowel. *Perception & Psychophysics, 25*, 457-465.

Mullenix, J. W. & Pisoni, D. B. (1990). Stimulus variability and processing dependencies in speech perception. *Perception & Psychophysics, 47*, 379-390.

Mullenix, J. W., Pisoni, D. B., & Martin, C. S. (1989). Some effects of talker variability on spoken word recognition. *Journal of the Acoustical Society of America, 85*, 365-378.

Nakatani, L. H. & Dukes, K. D. (1977). Locus of segmental cues for word juncture, *Journal of the Acoustical Society of America, 62*, 714-719.

Norris, D. (1982). Autonomous processes in comprehension: a reply to Marslen-Wilson and Tyler. *Cognition, 11*, 97-101.

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. (1993). Bottom-up connectionist models of "interaction". In G.

T. M. Altmann & R. Shillcock (Eds.), *Cognitive models of speech processing: the second Sperlonga meeting* (pp. 211-234). Hillsdale, NJ: Erlbaum.

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

Norris, D., McQueen, J.M.,& Cutler, A. (1995). Competition and segmentation in spoken word recognition. *Journal of Experimental Psychology: Learning, Memory & Cognition, 21*, 1209-1228.

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

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.

Nunn, A. M. (1998). *Dutch orthography: A systematic investigation of the spelling of Dutch words.* The Hague: Holland Academic Graphics.

Ooijen, B. van (1996). Vowel mutability and lexical selection in English: Evidence from a word reconstruction task. *Memory & Cognition, 24*, 573-583.

Otake, T., Hatano, G., Cutler, A., & Mehler, J. (1993). Mora or syllable? Speech segmentation in Japanese. *Journal of Memory and Language, 32*, 58-278.

Pisoni, D. B. & Luce, P. A. (1987). Acoustic-phonetic representations in word recognition. *Cognition, 25*, 21-52.

Pitt, M. A., & McQueen, J. M. (1998). Is compensation for coarticulation mediated by the lexicon? *Journal of Memory and Language, 39*, 347-370.

Pitt, M. A., & Samuel, A. G. (1995). Lexical and sublexical feedback in auditory word recognition. *Cognitive Psychology, 29*, 149-188.

Plaut, D., & Shallice, T. (1993). Deep dyslexia: A case study of connectionist neuropsychology. *Cognitive Neuropsychology, 10*, 377-500.

Pols, L. C. W. (1983). Three-mode principal component analysis of confusion matrices, based on the identification of Dutch consonants, under various conditions of noise and reverberation. *Speech Communication, 2*, 275-293.

Prager, R. W., Harrison, T. D. & Fallside, F. (1986). Boltzmann machines for speech recognition. *Computer Speech & Language, 1*, 3-27.

Quené, H. (1989). *The influence of acoustic-phonetic word boundary markers on perceived word segmentation in Dutch.* Unpublished doctoral dissertation, University of Utrecht, The Netherlands.

Quené, H. (1992). Integration of acoustic-phonetic cues in word segmentation. In M. E. H. Schouten (Ed.), *The Auditory Processing of Speech: from Sounds to Words* (pp. 349-355). Berlin: Mouton de Gruyter.

Quené, H. (1993). Segment durations and accent as cues to word segmentation in Dutch. *Journal of the Acoustical Society of America, 94*, 2027-2035.

Radeau, M., Morais, J., & Dewier, A. (1989). Phonological priming in spoken word recognition. *Memory & Cognition, 17*, 525-535.

Repp, B.,& Liberman, A. (1987). Phonetic category boundaries are flexible. In: S. Harnad (Ed.), *Categorical perception: the groundwork of cognition* (pp. 89-112). Cambridge, Cambridge University Press.

Roach, P., Knowles, G., Varadi, T., & Arnfield, S. (1993). MARSEC: A machine-readable spoken English corpus. *Journal of the International Phonetic Association, 23*, 47-53.

Saffran, J. R., Aslin, R. N., & Newport, E. L. (1996). Statistical learning by 8-month-olds. *Science, 274*, 1926-1928.

Saffran, J. R., Newport, E. L., & Aslin, R. N. (1996). Word segmentation: The role of distributional cues. *Journal of Memory and Language, 35*, 606-621.

Samuel, A. (1981). Phonemic restoration: Insights from a new methodology. *Journal of Experimental Psychology: General, 110*, 474-494.

Shillcock, R. C. (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.

Slowiaczek, L. M., & Hamburger, M. B. (1992). Prelexical facilitation and lexical interference in auditory word recognition, *Journal of Experimental Psychology: Learning, Memory and Cognition, 18*, 1239-1250.

Streeter, L. A., & Nigro, G. N. (1979). The role of medial consonant transitions in word perception. *Journal of the Acoustical Society of America, 65*, 1533-1541.

Suomi, K., McQueen, J. M., & Cutler, A. (1997). Vowel harmony and speech segmentation in Finnish. *Journal of Memory and Language, 36*, 422-444.

Taft, M. (1986). Lexical access codes in visual and auditory word recognition. *Language and Cognitive Processes, 1*, 297-308.

Taft, M., & Hambly, G. (1986). Exploring the cohort model of spoken word recognition. *Cognition, 22*, 259-282.

Turk, A. E., Jusczyk, P. W., & Gerken, L. A. (1995). Do English-learning infants use syllable weight to determine stress? *Language and Speech, 38,* 143-158.

Tyler, L. K. & Wessels, J. (1983). Quantifying contextual contributions to word-recognition processes. *Perception & Psychophysics, 34,* 409-420.

Tyler, L. K. & Wessels, J. (1985). Is gating an on-line task? Evidence from naming latency data. *Perception & Psychophysics, 38,* 217-222.

Vitevitch, M. S. & Luce, P. A. (1998). When words compete: Levels of processing in spoken word recognition. *Psychological Science, 9,* 325-329.

Vitevitch, M. S. & Luce, P. A. (1999). Probabilistic phonotactics and neighborhood activation in spoken word recognition. *Journal of Memory and Language. 40,* 374-408.

Vitevitch, M. S., Luce, P. A., Charles-Luce, J., & Kemmerer, D. (1997). Phonotactics and syllable stress: Implications for the processing of spoken nonsense words. *Language and Speech, 40,* 47-62.

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

Vroomen, J., Tuomainen, J., & de Gelder, B. (1998). The roles of word stress and vowel harmony in speech segmentation. *Journal of Memory and Language, 38,* 133-149.

Vroomen, J., van Zon, M., & de Gelder, B. (1996). Cues to speech segmentation: Evidence from juncture misperception and word spotting. *Memory & Cognition, 24,* 744-755.

Whalen, D. H. (1984). Subcategorical phonetic mismatches slow phonetic judgments. *Perception & Psychophysics, 35,* 49-64.

Whalen, D. H. (1991). Subcategorical phonetic mismatches and lexical access. *Perception & Psychophysics, 50,* 351-360.

Wurm, L. H., & Samuel, A. G. (1997). Lexical inhibition and attentional allocation during speech perception: Evidence from phoneme monitoring. *Journal of Memory and Language, 36,* 165-187.

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

Zwitserlood, P. (1996). Form priming. *Language and Cognitive Processes, 11,* 589-596.

APPENDICES

APPENDIX 3-A

Low and High Probability CV and VC sequences

Low Probability CV				
CV (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
bɛː	37	694	22	331
bɒː	28	145	6	18
dɑu	112	532	41	407
fɶy	95	796	40	232
fɛː	36	1434	6	133
fɒː	13	238	1	4
føː	65	2538	26	227
gɑ	36	264	16	33
gaː	60	463	4	14
geː	11	82	5	58
gɛ	29	212	14	210
giː	46	145	28	89
gɪ	5	0	5	0
gɛi	2	8	2	8
gau	2	16	2	16
gɔ	38	116	23	52
goː	75	594	33	228
guː	56	708	32	127
gə	47	994	16	330
hɛː	15	47	5	0
jiː	134	1909	1	0
jɒː	2	0	2	0
kɛː	45	865	3	91
kɒː	60	606	1	159

Low Probability CV

CV (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
lɒ:	18	227	6	199
mœ:	3	0	3	0
nœy	20	410	5	37
nɛ:	158	3479	1	0
pɛ:	42	581	9	115
rɛ:	43	1875	2	6
ʃɪ	87	1276	50	328
ʃɛi	12	101	10	77
ʃau	94	423	23	291
ʃu:	171	1077	42	45
ʃɛ:	28	207	4	31
sɛ:	74	2055	12	1210
ʃʉ	33	342	20	139
ʃɒ:	4	35	3	31
tɒ:	7	57	2	4
vɛ:	14	134	1	0
vø:	55	1468	25	155
wœy	62	1139	16	1003
wɒ:	9	74	9	74
wœ:	2	0	2	0
xœy	21	89	21	89
xy:	12	97	8	94
ʒɑ	69	3151	28	451
ʒɛ	66	1059	14	593
ʒy:	21	325	20	316
ʒɛ:	14	119	1	55
zɒ:	50	771	18	481
zø:	139	1869	48	888

Low Probability CV				
CV (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
ʒøː	34	55	7	30
dʒɑ	4	21	4	21
dʒaː	9	15	9	15
dʒeː	3	3	2	3
dʒɛ	53	346	49	343
dʒiː	13	619	9	617
dʒɪ	15	375	14	375
dʒɔ	28	210	27	210
dʒoː	2	36	2	36
dʒuː	4	37	4	37
dʒʉ	20	716	18	714

High Probability CV				
CV (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
bə	18591	909415	10445	724616
deː	5393	276296	1883	185681
dɪ	4032	232380	1157	151212
diː	5689	636230	1291	531958
doː	5291	278261	3940	247514
haː	2987	273898	1110	236406
hɑ	4740	397997	2374	331997
hɛ	3156	432772	2095	418900
hoː	3237	176845	1708	136522
kɑ	5332	195448	2604	170274
koː	4829	163525	2041	84052
kɔ	8768	324387	4753	261187

High Probability CV				
CV (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
la:	4616	185646	910	125607
lɑ	4779	270609	1833	141028
le:	6293	317923	2087	148139
lɪ	6368	251190	950	86070
li:	7105	214324	922	79847
mɑ	4795	199528	1958	119415
ma:	6634	596590	1470	436459
mɛ	3879	637378	1083	553611
me:	5544	323452	2318	235593
mɪ	3752	185706	1689	107577
mo:	3404	136444	1042	98454
na:	6134	486503	2234	359291
ne:	5906	213163	1316	91605
re:	6934	163709	2730	110472
tɛ	3620	189699	1233	135220
te:	8235	261025	1661	135920
tə	51488	2233632	1288	600565
vɑ	2889	1373347	1305	1260672
vɔ	3246	237672	1310	156925
vo:	3947	536404	3070	488136
və	30250	1347902	14584	608746
wɑ	2876	649005	1367	570236
wa:	4843	388103	1664	320347
we:	3898	330998	1535	208026
wɛ	7007	516055	3263	402684
wɪ	3750	204473	1406	147532
xə	21111	1218892	14460	1054389

Low Probability VC

VC (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
ɑʃ	23	263	4	240
ɛʃ	33	237	14	130
iːŋ	17	36	1	0
ɪʃ	3	140	1	4
iːʃ	6	37	4	34
ɛij	1	91	1	91
œyj	2	6	2	6
auk	17	5	3	0
aun	101	1302	6	304
aur	2	31	2	31
ɔj	73	1066	19	737
ɔʃ	5	95	3	95
oːw	50	316	18	247
uːŋ	55	148	3	11
uːʃ	28	837	14	791
yːl	22	140	11	140
yːm	43	570	26	540
yːn	14	204	7	188
yːp	3	15	2	6
yːʃ	2	86	2	86
yːx	9	226	6	213
ɛːl	4	0	2	0
ɛːm	32	354	17	261
ɛːn	46	396	4	330
ɛːs	33	201	29	191
ɛːt	7	38	3	24

Low Probability VC				
VC (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
ʉŋ	137	2346	1	22
ʉʃ	3	17	1	17
ɒ:k	9	100	2	84
ɒ:l	24	256	14	239
ɒ:r	76	870	25	651
ɒ:s	8	13	8	13
œ:r	16	19	1	19
ø:f	9	203	5	186
ø:m	36	265	5	2
ø:p	28	353	3	268
ø:x	33	899	15	502
əf	4	73	4	73
əp	18	47	6	47
əʃ	12	75	5	75

High Probability VC				
VC (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
a:l	3344	153371	1492	115459
ɑl	3641	356135	628	254085
ɑn	10224	1919413	511	1534669
a:n	5930	593872	1005	449067
a:r	7103	1298569	1672	963701
ɑs	4504	392138	821	345047
ɑt	2584	1116235	1106	1094084
a:t	3879	227837	2013	165621
ɑx	6156	291178	708	130933
e:l	2636	221246	732	174553

High Probability VC				
VC (IPA)	Syllable Type	Syllable Token	Word Type	Word Token
ɛl	4326	341771	888	179809
eːn	2845	327794	584	270680
eːr	13623	460916	4183	293480
ɛt	2763	494080	1273	466421
ɪŋ	11879	576211	7212	514911
ɛin	2647	738563	868	678318
ɛit	6071	340907	5188	326466
œyt	6875	334704	705	201574
ɑuw	2069	202498	663	182385
ɔn	12153	601176	629	91711
ək	2343	310412	909	267194
əl	13057	194333	4747	119733
ər	52699	1954686	15477	856487
əs	8703	159207	6984	135687
əx	6693	287683	3167	238937

Appendix 3-B

Target-bearing stimuli used in Experiments 3.1,3.2, & 3.3

Initial Targets; Experiment 3.1+3.2			Final Targets, Experiment 3.1+3.3		
High	Low	Translation	High	Low	Translation
bok-dif	bok-dous	goat	beel-deuk	buul-deuk	dent
ham-dil	ham-douf	ham	seel-douche	suul-douche	shower
kap-dirm	kap-doum	cap	kweel-rib	wuul-rib	rib
long-dir	long-doup	lung	vreel-rang	muul-rang	rank
zaag-dils	zaag-douk	saw	breel-teug	tuul-teug	sip
big-moof	big-nuif	piglet	gan-ram	goum-ram	ram
duif-mool	duif-nuim	pigeon	lan-riem	loun-riem	belt
pet-moon	pet-nuip	hat	tan-lap	toun-lap	rag
gif-moop	gif-nuik	poison	stan-leus	woun-leus	slogan
juf-mools	juf-nuig	nanny	nan-lood	noun-lood	lead
kiem-vam	kiem-guum	seed	fag-pauw	feug-pauw	peacock
duim-varf	duim-guuf	thumb	kag-put	veug-put	well
dam-vap	dam-guup	dam	tag-kuif	peug-kuif	forelock
rum-vag	rum-guug	rum	zwag-kus	neug-kus	kiss
kam-vams	kam-guun	comb	knag-mug	weug-mug	mosquito
boon-waam	boon-wuim	bean	kaat-goot	gouk-goot	gutter
deun-waaf	deun-wuig	tune	taat-geul	louk-geul	gully
map-waamt	map-wuip	folder	zwaat-mus	wouk-mus	sparrow
nul-waaft	nul-wuik	zero	knaat-non	stouk-non	nun
tor-waapt	tor-wuin	beetle	snaat-gips	nouk-gips	plaster
bil-wam	bil-zeuk	buttock	get-fuik	stuup-fuik	fyke
bon-wap	bon-zeut	receipt	ket-mouw	duup-mouw	sleeve
duin-wan	duin-zeun	dune	tet-mol	puup-mol	mole
pap-wamt	pap-zeus	mash	klet-nut	luup-nut	utility
lam-waps	lam-zeunt	lamb	stet-nok	muup-nok	ridge

Initial Targets; Experiment 3.1+3.2			Final Targets, Experiment 3.1+3.3		
High	Low	Translation	High	Low	Translation
peuk-wif	peuk-guif	butt	tat-kok	deup-kok	cook
pil-wing	pil-guik	pill	snat-kies	geup-kies	molar
lol-wimt	lol-guip	fun	brat-cape	leup-cape	cape
jeuk-wift	jeuk-guin	itch	grat-nis	meup-nis	niche
sap-wifs	sap-guim	juice	smat-jack	neup-jack	jacket
sein-voof	sein-veur	signal	fon-lus	duun-lus	loop
tang-vook	tang-veus	pliers	gon-loep	buun-loep	lens
mam-vool	mam-veul	mom	lon-rit	kuun-rit	ride
ring-voom	ring-veuk	ring	mon-lek	muun-lek	leak
loon-voon	loon-veut	wages	kron-roos	vuun-roos	rose
pol-vof	pol-feur	clump	stein-rek	deum-rek	rack
tol-vom	tol-feul	toll	nein-wol	geum-wol	wool
zang-von	zang-feuk	song	slein-wok	leum-wok	wok
lak-voft	lak-feus	varnish	smein-doek	meum-doek	towel
kuip-voms	kuip-feukt	barrel	krein-bek	neum-bek	snout

APPENDIX 3-C

Target-bearing stimuli used in Experiments 3.4 & 3.5

High Probability	Target Translation	Low Probability	Target Translation
Initial Targets, Experiment 3.4			
drum-baaf	drum	creme-baag	cream
dief-beem	thief	sleuf-beep	groove
pus-bif	pus	hasj-bim	hash
spies-bouf	skewer	creche-boug	daycare
kroos-daaf	duckweed	gleuf-daam	slot
pluis-deup	fluff	vleug-deut	breath
vlas-doum	flax	teug-dous	sip
slag-duip	beat	spuug-duis	spit
speen-fiek	teat	clown-fiet	clown
kluis-foek	safe	douche-foes	shower
pauw-gam	peacock	kieuw-gan	gills
koek-heem	cake	truc-heeg	trick
kap-juif	cap	heup-juik	hip
muis-ruuf	mouse	zeug-ruus	sow
kwaal-suum	disease	beul-suun	hangman
Final Targets, Experiment 3.5			
kaaf-zeem	chammy	taaf-zeur	nag
baam-dis	dish	daam-jazz	jazz
beeg-dok	dock	peeg-junk	junkie
jeel-vaas	vase	weel-vouw	fold
beem-dot	dowry	deem-dauw	dew
deep-wees	orphan	heep-woud	forest
teuf-wig	wedge	veuf-jicht	gout
peum-vaat	dishes	veum-fout	error

High Probability	Target Translation	Low Probability	Target Translation
meun-zeef	sieve	neun-sjerp	sash
seup-vod	rag	veup-fohn	hairdryer
deus-wal	shore	meus-jack	jacket
fiem-gal	gall	siem-geul	gully
hien-galg	gallows	pien-geur	smell
bif-dam	dam	tif-jeans	jeans
goem-vat	barrel	loem-fuik	fyke
boof-gans	goose	toof-shirt	shirt
bouk-ven	puddle	fouk-cheque	cheque
boum-vork	fork	foum-zeug	sow
kuuf-gast	guest	puuf-sjeik	sheik
guuk-zeep	soap	suuk-chef	boss
buum-vel	skin	puum-saus	sauce

APPENDIX 4-A

Materials Experiments 4.1-4.4

Target-bearing triplets: Final targets, initial mismatches				
Word Onset (completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedded	Translation Embedding
syn.thee(se)	gynthee	bynthee	tea	synthesis
cho.les(terol)	foles	boles	lesson	cholesterol
cy.nis(me)	gynis	bynis	niche	cynism
far.ma(ceutisch)	garma	larma	mom	farmaceutical
fi.jas(co)	gijas	nijas	jacket	fiasco
ga.lak(tisch)	falak	talak	varnish	galactic
gor.dij(nen)	fordij	mordij	thigh	curtains
sa.dis(me)	fadis	badis	dish	sadism
se.lek(tie)	felek	melek	leak	selection
se.mes(ter)	femes	lemes	knife	semester
ser.pen(tine)	gerpen	berpen	pen	streamer
si.ree(ne)	giree	biree	roe	siren
sym.po(sium)	fympo	bympo	pot	symposium
sy.nop(sis)	ginop	linop	stud	synopsis
va.kan(tie)	zakan	takan	jug	holiday
ver.heer(lijken)	zerheer	kerheer	gentleman	glorify
ver.hel(deren)	zerhel	kerhel	hell	clarify
ver.koe(ling)	zerkoe	kerkoe	cow	cooling
ju.wee(len)	luwee	puwee	contraction	jewel
ra.bar(ber)	wabar	kabar	bar	rhubarb
ro.ton(de)	woton	koton	cask	roundabout
re.dak(tie)	wedak	pedak	roof	editorial board
re.la(tie)	wela	kela	drawer	relation
re.por(tage)	wepor	fepor	poke	report

Target-bearing triplets: Final targets, initial mismatches				
Word Onset (completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedded	Translation Embedding
ru.moe(rig)	wumoe	pumoe	tired	noisy
mai.zee(na)	naizee	kaizee	sea	cornflour
mo.nar(chie)	nonar	gonar	fool	monarchy
ca.brie(olet)	tabrie	zabrie	brie	convertible
ca.pa(citeit)	tapa	zapa	dad	capacity
ka.das(ter)	tadas	vadas	tie	land register
par.man(tig)	tarman	zarman	man	jaunty
pi.roe(ette)	kiroe	niroe	rod	pirouette
ta.veer(ne)	kaveer	vaveer	feather	tavern

Matched filler triplets			
Word Onset(completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedding
certi(ficaat)	ferti	merti	certificate
circu(leren)	gircu	bircu	circulate
fanfa(re)	ganfa	nanfa	brass band
fili(aal)	gili	nili	branch
gori(lla)	fori	mori	gorilla
sabo(tage)	fabo	babo	sabotage
sensa(tie)	fensa	bensa	sensation
simpli(ficatie)	fimpli	bimpli	simplification
simu(latie)	gimu	limu	simulation
situ(atie)	fitu	mitu	situation
soli(dariteit)	goli	joli	solidarity
subli(meren)	gubli	lubli	sublimate
sugges(tie)	fugges	bugges	suggestion
symfo(nie)	fymfo	bymfo	symphony
verbeel(den)	zerbeel	kerbeel	imagine

Matched filler triplets			
Word Onset(completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedding
verbor(gen)	zerbor	kerbor	hidden
verduis(teren)	zerduis	kerduis	darken
virtu(oos)	zirtu	tirtu	virtuoso
jovi(aal)	lovi	tovi	jovial
jubi(leum)	lubi	pubi	jubilee
justi(tie)	wusti	tusti	justice
recep(tie)	wecep	fecep	reception
recher(che)	wecher	pecher	detective
regre(ssie)	wegre	kegre	regression
rota(tie)	wota	kota	rotation
migra(tie)	nigra	kigra	migration
miljar(dair)	niljar	siljar	biljonaire
cumu(latief)	tumu	zumu	cumulative
peda(goog)	keda	leda	educationist
pio(nier)	tio	nio	pioneer
publi(citeit)	tubli	zubli	publicity
timi(de)	kimi	vimi	timid
turbu(lentie)	purbu	murbu	turbulence

APPENDIX 5-A

Materials Experiments 5.1-5.3

Target-bearing triplets: Initial targets, final mismatches

Word Onset (completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedded	Translation Embedding	
ar.ties(ten)	artieg	artiel	sledge	artist	*
as.fal(teren)	asfaw	asfam	ashes	asphalt	
bar.bec(ue)	barbep	barbem	bar	barbecue	
bil.jar(ten)	biljal	biljang	bottom	billiards	*
boer.gon(je)	boergong	boergof	farmer	Burgundy	
kul.min(atie)	kulming	kulmiek	rubbish	culminate	
fee.noom(een)	feenoon	feenook	fairy	phenomenon	
gal.vaan(iseren)	galvaam	galvaaf	gall	galvanise	*
ham.bur(ger)	hambuj	hambup	ham	hamburger	*
hek.tool(iter)	hektoor	hektoop	gate	hectoliter	*
hen.dic(ap)	hendiet	hendiem	hen	handicap	*
kam.piej(oen)	kampiew	kampieg	comb	champion	*
kan.del(aar)	kandew	kandem	jug	candelabra	*
kap.suul(e)	kapsuuw	kapsuuk	cap	capsule	*
klep.toom(aan)	kleptoong	kleptook	valve	kleptomaniac	*
kom.pliem(ent)	komplien	komplies	bowl	compliment	*
la.wien(e)	lawiem	lawief	drawer	avalanche	
lek.siek(on)	leksiet	leksiem	leak	lexicon	*
ma.jes(teit)	majeg	majek	mum	majectic	
man.jif(iek)	manjis	manjik	man	magnificent	
nok.tur(ne)	noktul	noktup	ridge	nocturne	*
pa.gien(a)	pagiem	pagiep	dad	page	
po.puul(air)	popuuw	popuuk	pot	popular	

Note. The 18 target-bearing triplets used in Experiment 5.4 are marked with an asterisk.

Target-bearing triplets: Initial targets, final mismatches

Word Onset (completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedded	Translation Embedding	
por.tuug(al)	portuuf	portuup	thrust	Portugal	
rang.cuun(eus)	rangcuum	rangcuuk	rank	vindictive	
ree.ser(vaat)	reesel	reesep	doe	reservate	
rek.tief(icatie)	rektieg	rektiek	rack	rectification	
sla.ver(nij)	slavel	slavep	lettuce	slavery	*
som.breer(o)	sombreew	sombrees	sum	sombrero	*
tak.soon(omie)	taksoom	taksoof	branch	taxonomy	*
tor.tiel(la)	tortier	tortief	beetle	tortilla	
trom.bon(e)	trombom	trombop	drum	trombone	*
wak.sien(elichtje)	waksiem	waksief	ice-hole	candle	*

Note. The 18 target-bearing triplets used in Experiment 5.4 are marked with an asterisk.

Matched filler triplets

Word Onset (completion)	Minimal Mismatch	Maximal mismatch	Translation embedding
certif(icaat)	certig	certik	certificate
circul(eren)	circuw	circuk	circulate
fanfar(e)	fanfaj	fanfap	brass band
filij(aal)	filiw	filik	branch
sensat(ie)	sensap	sensaf	sensation
simplif(icatie)	simplig	simplim	simplification
simul(atie)	simuw	simuk	simulation
situw(atie)	situl	situn	situation
sugges(tie)	suggef	suggel	suggestion
symfon(ie)	symfoom	symfook	symphony
verbeel(den)	verbeew	verbeep	imagine
verbor(gen)	verbow	verbog	hidden
verduis(teren)	verduig	verduip	obscure

Matched filler triplets			
Word Onset (completion)	Minimal Mismatch	Maximal mismatch	Translation embedding
jubil(eum)	jubiw	jubig	jubilee
justit(ie)	justik	justig	justice
recep(tie)	recek	receg	reception
recher(che)	rechew	rechep	detective
regres(sie)	regref	regrep	regression
rotat(ie)	rotap	rotam	rotation
migrat(ie)	migrap	migraw	migration
miljar(dair)	miljaw	miljam	billionaire
cumul(atief)	cumuw	cumup	cumulative
pedag(oog)	pedaf	pedar	pedagogue
public(iteit)	publif	publin	publicity
turbul(entie)	turbuw	turbum	turbulence
absol(utie)	absow	absop	absolution
bonif(icatie)	bonig	bonip	bonification
chirur(gie)	chirul	chirup	surgery
faillis(ement)	faillif	faillim	bankruptcy
lantar(en)	lantaw	lantam	street light
molec(uul)	molep	molef	molecule
parlem(ent)	parlen	parler	parliament
stimul(eren)	stimuw	stimuk	stimulate

APPENDIX 5-B

Additional Experimental Materials Experiment 5.4

Target-bearing triplets: 18 Final targets, Initial mismatches				
Word Onset (completion)	Minimal Mismatch	Maximal Mismatch	Translation Embedded	Translation Embedding
cy.nis(me)	gynis	bynis	niche	cynicism
fi.jas(co)	gijas	nijas	jacket	fiasco
ga.lak(tisch)	falak	talak	varnish	galactic
gor.dij(nen)	fordij	mordij	thigh	curtains
se.mes(ter)	femes	lemes	knife	semester
sy.nop(sis)	ginop	linop	stud	synopsis
syn.thee(se)	gynthee	bynthee	tea	synthesis
ver.hel(deren)	zerhel	kerhel	hell	clarify
ra.bar(ber)	wabar	kabar	bar	rhubarb
ro.ton(de)	woton	koton	cask	roundabout
re.dak(tie)	wedak	pedak	roof	editorial board
re.la(tie)	wela	kela	drawer	relation
ru.moe(rig)	wumoe	pumoe	tired	noisy
mai.zee(na)	naizee	kaizee	sea	cornflour
mo.nar(chie)	nonar	gonar	fool	monarchy
ka.das(ter)	tadas	vadas	tie	land register
par.man(tig)	tarman	zarman	man	jaunty
ta.veer(ne)	kaveer	vaveer	feather	tavern

SAMENVATTING

VAN SPRAAK NAAR WOORDEN

Hoe komen wij van spraak naar woorden, van akoestisch signaal naar betekenis? Voordat gesproken taal kan worden herkend als een reeks van woorden die een betekenisvolle boodschap overdraagt van spreker naar luisteraar, moeten er twee belangrijke problemen worden opgelost door de luisteraar: het segmentatieprobleem en het variatieprobleem.

Het segmentatie probleem

Hoe weten wij waar het ene woord eindigt en het volgende woord begint? Het spraaksignaal is continu. In geschreven taal geven spaties op een eenduidige manier de grenzen tussen de woorden aan; in gesproken taal bestaat zo'n duidelijke markering van de woordgrenzen niet. Het gemak waarmee we woorden in onze moedertaal waarnemen is dan ook misleidend. Dit wordt op een pijnlijke manier duidelijk wanneer je een taal probeert te begrijpen die je niet spreekt. Bijvoorbeeld wanneer een ééntalige Sloveense bezoekster aan het Kröller-Müller museum haar groep is kwijtgeraakt en in vloeiend Sloveens probeert uit te leggen wat er precies is gebeurd: je hoort een brij van klanken zonder de boodschap te begrijpen. Om te begrijpen wat er gezegd wordt moet de luisteraar de gesproken uitingen segmenteren in woorden. Hoe lost de luisteraar dit segmentatieprobleem op? In het eerste gedeelte van dit onderzoek is gekeken in hoeverre luisteraars gebruik maken van de waarschijnlijkheid dat bepaalde klanken op een bepaalde plaats binnen een woord voorkomen.

Het variatieprobleem

De vorm waarin de klanken van gesproken taal het oor van de luisteraar bereiken, is enorm gevarieerd door een groot aantal factoren. Sommige sprekers hebben een lage stem, andere sprekers hebben een hoge stem; sommige sprekers hebben een Achterhoeks dialect, weer andere praten Betuws; soms zijn ze schor; dan weer praten zij snel omdat zij haast hebben of langzaam als zij moe zijn; je praat en luistert in een kroeg met luid geroezemoes of in een discotheek met harde muziek. Niettemin is de menselijke spraakherkenning opmerkelijk stabiel. We begrijpen wat er wordt gezegd, ondanks de voortdurend veranderende luisteromstandigheden. Om flexibel te kunnen reageren op al deze veranderingen, zou je verwachten dat we woorden ook kunnen herkennen als ze niet helemaal goed zijn uitgesproken: hoe tolerant is gesproken woordherkenning als wat je hoort niet spoort[1] met een bestaand (Nederlands) woord? Het klassieke voorbeeld is gegeven door Norris (1982): de luisteraar zou moeten weten wat de dronken man bedoelt als hij vraagt om een *sjigaret*. Het onderwerp van het tweede gedeelte van dit onderzoek is de gevoeligheid (of ongevoeligheid) van woordherkenning met betrekking tot misleidende informatie in de gesproken taal.

De 'woord spotting'-taak

De 'woord spotting'-taak wordt binnen de taalpsychologie veel gebruikt om woordherkenning in continue spraak te bestuderen. Bij een 'woord spotting'-taak (zie ook Figuur 1-1 in Hoofstuk 1) luisteren de proefpersonen naar een reeks van onzinwoorden. Het is de bedoeling dat de proefpersonen woorden *ontdekken* die aan het begin, of aan het eind van de onzinwoorden zijn ingebed. Met een druk op een knop moeten de proefpersonen zo snel en precies mogelijk aangeven dat ze een bestaand woord hebben gehoord dat verstopt zat in het onzinwoord, en vervolgens moeten zij dit bestaande woord hardop uitspreken, zodat er gecontroleerd kan worden of zij inderdaad het juiste woord hebben ontdekt. Als de proefpersonen geen bestaand woord hebben herkend in wat ze gehoord hebben, dan hoeven zij niet te reageren. Bijvoorbeeld, als je *publi*, *deemdeep* of *luutvoem* hoort, dan reageer je niet. Hoor je *boomdouf*, *choles*, of *hambur*, dan reageer je wel, omdat *boom*, *les* en *ham* bestaande Nederlandse woorden zijn. Deze 'woord spotting'-taak is gebruikt in experimenten die gerapporteerd worden in Hoofdstuk 2, 3, 4 en 5.

De 'identiteits priming'-taak

Bij een 'identiteits priming'-taak (zie ook Figuur 1-2 in Hoofdstuk 1) luisteren de proefpersonen naar een reeks van talige geluidsfragmenten. Deze geluidsfragmenten kunnen delen van woorden zijn, zoals losse klanken en lettergrepen, of langere stukken gesproken taal, zoals woorden en zelfs

---

[1] *De term 'niet sporen' als Nederlandse verwoording van het Engelse 'mismatching' werd gesuggereerd door Albert van der Hem.*

complete zinnen. Op een bepaald moment (hetzij tijdens of net na het horen van een geluidsfragment) verschijnt er op een beeldscherm voor de proefpersoon een reeks letters. De proefpersoon wordt gevraagd om te bepalen of deze reeks letters een bestaand woord vormt of niet en dit aan te geven door middel van het indrukken van de juiste knop op een knoppenkast (er is een groene JA knop en een rode NEE knop). Een aantal van de letterreeksen komt in vorm overeen met het geluidsfragment dat de proefpersonen net hebben gehoord, bijvoorbeeld, als je *hambur* hoort en je ziet vervolgens de letterreeks HAM. De snelheid waarmee mensen bepalen of een bepaalde letterreeks een woord vormt net nadat ze een geluidsfragment hebben gehoord waar dat woord in voorkwam, en de hoeveelheid fouten die zij daarbij maken, wordt vergeleken met de snelheid en precisie in gevallen waar er geen overeenkomst is tussen wat de proefpersoon hoort en wat de proefpersoon ziet, bijvoorbeeld, als je *rangcuun* hoort en vervolgens moet beslissen of de letterreeks HAM een bestaand woord is of niet. Deze 'identiteits priming'-taak is gebruikt in experimenten die worden beschreven in Hoofdstuk 4 en 5.

KLANKWAARSCHIJNLIJKHEID IN SEGMENTATIE

In het eerste experiment (Hoofdstuk 2) is gekeken of luisteraars bij het in woorden opsplitsen van de continue gesproken taal gebruik kunnen maken van de waarschijnlijkheid dat bepaalde klanken op een bepaalde plaats voorkomen. Het horen van klanken die vaak voorkomen aan het begin van een woord (bijvoorbeeldde klank *h* in *hemd*, *honger*, *hanggeranium*, etcetera) zou het herkennen van het voorafgaande woord kunnen vergemakkelijken omdat de *h* duidt op het begin van iets; het horen van klanken die vaak het eind vormen van een woord (bijvoorbeeld de klank *ng* in *long*, *nijptang*, *doorzonwoning*, etcetera) zou kunnen helpen bij het herkennen van het daaropvolgende woord omdat *ng* wijst op het eind van iets. Een 'woord spotting'-taak is gebruikt om te kijken of luisteraars gebruik kunnen maken van de relatieve waarschijnlijkheid van een zestal klankcombinaties. De zes combinaties van klanken die getest werden, kunnen allemaal zowel aan het begin als aan het eind voorkomen van Nederlandse woorden. Het verschil zit daarin dat sommige klanken vaker aan het begin van woorden voorkomen en andere klanken vaker aan het eind van een woord. Bijvoorbeeld, de klankcombinatie ***sp*** komt vaker voor aan het begin van een woord dan aan het eind van een woord in het Nederlands. De klankcombinatie ***ks***, daarentegen, komt vaker voor aan het eind van een woord. Proefpersonen luisterden naar onzinwoorden als *bloem**spuum***, *bloem**ks**uum*, *duu**sp**vriend* en *duu**ks**vriend*. Hierbij was de verwachting dat de woorden *bloem* en *vriend* het gemakkelijkst zouden kunnen worden ontdekt in *bloem**sp**uum* en in *duu**ks**vriend* omdat daar de klankcombinatie op de meest waarschijnlijke positie werd gehoord. De resultaten leverden geen sterk bewijs op dat luisteraars dit soort informatie gebruiken bij het segmenteren van spraak. Misschien is dat niet verwonderlijk als je je bedenkt dat, hoewel de klankcombinatie ***sp*** veel vaker aan het begin van een woord zit, er toch nog 52 woorden in het Nederlands zijn die eindigen met de klankcombinatie ***sp***. De relatieve frequentie die werd gebruikt voor het bepalen van de waarschijnlijkheid dat een bepaalde klank op een bepaalde plaats voorkomt is daarom misschien een te onbetrouwbare indicatie van de waarschijnlijkheid van een woordgrens. Bovendien speelt de klank *s* een bijzondere rol in complexe combinaties van medeklinkers. In een serie van zeven vervolgexperimenten zijn daarom de absolute frequenties van klinker-medeklinker en medeklinker-klinker combinaties onderzocht.

In Hoofdstuk 3 worden de resultaten gerapporteerd van vijf experimenten waarbij de 'woord spotting'-taak is gebruikt om te testen of de absolute frequentie van medeklinker-klinker (CV) combinaties aan het begin van woorden of klinker-medeklinkercombinaties (VC) aan het eind van woorden bijdraagt aan het herkennen van Nederlandse woorden in langere stukken spraak. Deze CV en VC combinaties kwamen in het geluidsfragment voor, hetzij net naast het ingebedde bestaande woord (zoals het hoogfrequente ***di*** in *boom**dif*** of het laagfrequente ***dou*** in *boom**dou**f*), hetzij als deel van het woord grenzend aan de context (zoals het hoogfrequente ***ap*** in *k**ap**juif* of het laagfrequente ***eup*** in ***heupjuik***). Hierbij werd verondersteld dat het makkelijker zou zijn om woorden te ontdekken als de klankcombinatie die grensde aan de woordgrens vaak voorkomt op die positie in een woord. Twee extra controle-experimenten zijn uitgevoerd om te bepalen of de klankwaarschijnlijkheid ook gevolgen had bij het bepalen of een geluidsfragment een bestaand Nederlands woord vormt ('lexicale decisie'-taak). Deze laatste twee experimenten waren bedoeld om de mogelijkheid uit te sluiten dat de effecten in de voorgaande experimenten het gevolg waren van het beslissen of een bepaald stuk spraak een bestaand woord vormt of niet, in plaats van het veronderstelde gevolg van de klankwaarschijnlijkheid bij het vinden van de grenzen tussen woorden.

Uit de resultaten van Hoofdstuk 3 kan worden geconcludeerd dat luisteraars inderdaad kennis over de waarschijnlijkheid dat een bepaalde klank op een bepaalde plek voorkomt, gebruiken bij de segmentatie van gesproken taal. Het gebruik van deze informatie lijkt echter op een tweetal manieren beperkt te zijn. Ten eerste beïnvloedt de waarschijnlijkheid van een klankreeks de segmentatie meer wanneer de reeks voorkomt binnen het te ontdekken woord. Ten tweede lijkt informatie omtrent de waarschijnlijkheid dat een klankcombinatie voorkomt aan het begin van een woord het meest effectief te worden gebruikt door de luisteraar. Dat het gebruik van deze kennis beperkt is, is niet verwonderlijk, omdat de aanwijzing die deze informatie geeft over de locatie van een woordgrens relatief zwak is. Een klankcombinatie die weinig voorkomt als het begin van een woord is toch nog steeds het begin van een aantal woorden (die woorden moeten we op de een of andere manier toch ook kunnen herkennen); een klankcombinatie die vaak voorkomt als het begin van een woord is waarschijnlijk, maar niet noodzakelijk het begin van een woord.

MISLEIDENDE INFORMATIE EN COMPETITIE

Er is een enorme variatie in de realisatie van spraakgeluiden. Niettemin is de menselijke spraakperceptie opmerkelijk stabiel. Hoe lossen luisteraars dit variatieprobleem op?

Omdat de luisteraar zich op de een of andere manier aan moet passen aan de alsmaar veranderende luisteromstandigheden zou je in het proces van woordherkenning een zekere flexibiliteit verwachten met betrekking tot niet helemaal correct uitgesproken woorden. In Hoofdstuk 4 en 5 zijn zes 'woord spotting'- en vier 'identiteits priming'-experimenten gebruikt om te onderzoeken hoe tolerant gesproken woordherkenning is in gevallen waar wat je hoort niet spoort met een woord, dat wil zeggen, in gevallen waar de informatie in het spraaksignaal niet overeenstemt met de informatie die is opgeslagen in het mentale lexicon. In het Nederlands kunnen spraakklanken op drie manieren verschillen: in de plaats van articulatie (bijvoorbeeld, *t* en *p*), in stemhebbendheid (bijvoorbeeld, *t* en *d*) en in de manier van articuleren (bijvoorbeeld, *t* en *s*). In deze experimenten werden verschillen ter grootte van één fonetische eigenschap, de plaats van articulatie, gecontrasteerd met verschillen in alledrie deze fonetische eigenschappen (klanken die zowel verschilden in plaats, stemhebbendheid en manier van articuleren (bijvoorbeeld, t en m). De gevoeligheid werd op een indirecte manier gemeten door te kijken naar de competitie tussen een woord, dat mogelijk in voldoende mate herkend zou kunnen worden ondanks de misleidende informatie in het spraaksignaal, en andere woordkandidaten. McQueen, Norris en Cutler (1994) hebben voor het Engels aangetoond dat het moeilijker is om een woord als *les* te ontdekken in een fragment als *choles* dan in een fragment als *boles*, omdat *choles* ook het begin vormt van het langere woord *cholesterol*. De woorden *les* en *cholesterol* wedijveren in dit geval om herkenning. Dit gegeven is in Hoofdstuk 4 en 5 gebruikt om het effect van misleidende informatie op gesproken woordherkenning te onderzoeken.

In vier experimenten (Hoofdstuk 4), kwam de misleidende informatie aan het begin van de fragmenten voor. De herkenning van *les* in het geval van een 'Woordbegin', bijvoorbeeld *choles* (het begin van *cholesterol*), werd vergeleken met een 'Minimale Misleiding', bijvoorbeeld *foles*, en met een 'Maximale Misleiding', bijvoorbeeld *boles*. In de andere zes experimenten (Hoofdstuk 5) werd het effect van misleidende informatie op een latere positie in het woord onderzocht. De herkenning van *ham* in *hambur* werd bijvoorbeeld vergeleken met de herkenning van *ham* in *hambuj* (Minimale Misleiding) en de herkenning van *ham* in *hambup* (Maximale Misleiding). Als het langere woord (*cholesterol* of *hamburger*) zelfs in het geval van misleidende informatie tot op zekere hoogte wordt herkend dan zou zich dat moeten uiten door competitie-effecten (vanwege competitie tussen *les* en *cholesterol* of tussen *ham* en *hamburger*).

Op grond van de resultaten in Hoofdstuk 4 en 5 samen kan geconcludeerd worden dat luisteraars tamelijk intolerant en inflexibel zijn met betrekking tot misleidende informatie. Dat is niet verwonderlijk als je bedenkt wat de uiterste consequentie is van dit soort flexibiliteit: we zouden constant dingen horen die niet worden gezegd en hallucineren in plaats van taal begrijpen. Bovendien tonen de resultaten ook aan dat kleine afwijkingen minder goed te onderscheiden zijn dan grote afwijkingen, ongeacht de positie van de misleidende informatie in het woord. In de algemene discussies van Hoofdstuk 4 en 5 worden een aantal factoren besproken, die van invloed kunnen zijn op het vinden van effecten van misleidende informatie in taalpsychologische experimenten. Ten eerste lijkt dit soort manipulaties erg ontvankelijk te zijn voor het gebruik van strategieën door de proefpersonen. Ten tweede is het belangrijk hoeveel informatie er in het geluidsfragment aanwezig is ter ondersteuning van de herkenning van het woord. Deze tweede factor wordt ook nog beïnvloed door een derde belangrijke factor: de luisteromstandigheden. Als je met iemand zit te praten in een rumoerige kroeg, dan ben je meer geneigd om flexibel om te gaan met de informatie die het oor bereikt, dan wanneer je in een geluidsdichte kabine zit te luisteren.

HET DILEMMA VAN DE EXPERIMENTATOR

De taalpsycholoog is geïnteresseerd in één van de grootste prestaties van de mens: het vermogen om taal te verwerven, te begrijpen en te spreken. Om meer te weten te komen over dit interessante onderwerp moet de onderzoeker de onderzoekstof opdelen in hapklare brokken en zich beperken (bijvoorbeeld, door zich te richten op specifieke onderwerpen als segmentatie en misleidende informatie in gesproken woorherkenning). Vervolgens draait de taalpsycholoog zorgvuldig gecontroleerde experimenten met proefpersonen van de Nintendo generatie (Zullen we een spelletje doen?) en met testmateriaal dat is geselecteerd om alle mogelijke verwarring met andere aspecten uit te sluiten. De resultaten zijn veelbelovend, maar wat vertellen ze ons nu werkelijk over taal in het wild? In een voorstudie in het veld, die werd uitgevoerd in de bar van Vergader- en Trainingscentrum Jonkerbosch in Nijmegen, werd tegen 25 proefpersonen (voor deelname beloond met bier) de volgende testzin uitgesproken:"maggikunzigaletjevajuhbietsuh?" De resultaten suggereren dat minimale afwijkingen mijn gezondheid nog ernstig schade kunnen toebrengen.

CURRICULUM VITAE

Arie van der Lugt werd geboren in Bennekom op 20 juli 1968. Na het behalen van zijn gymnasiumdiploma aan het Marnix College te Ede studeerde hij informatica en Wijsbegeerte van Wetenschap, Technologie en Samenleving aan de Universiteit Twente. Tijdens het laatste gedeelte van zijn studie werkte hij als stagiaire op het Nijmeegs Instituut voor Cognitie en Informatie in Nijmegen en vanaf maart 1994 als onderzoeksassistent aan het Max Planck Instituut voor Psycholinguïstiek (MPI) in Nijmegen. Na het behalen van zijn ingenieursdiploma in augustus 1994 bleef hij als onderzoeksassistent verbonden aan het MPI. In oktober 1995 werd hem een stipendium toegekend door de Max Planck Gesellschaft zur Förderung der Wissenschaften om promotieonderzoek te doen aan het MPI binnen de onderzoeksgroep 'Spoken Word Recognition'. Tijdens zijn academische opleiding (van 1986 tot 1999) was hij ook werkzaam als suppoost/gastheer in deeltijd bij het Kröller-Müller museum te Otterlo. Bovendien was hij van november 1998 tot oktober 1999 in dienst bij het Vergader- en Trainingscentrum Jonkerbosch in Nijmegen als nachtportier. Vanaf oktober 1999 is hij als postdoctoral research fellow verbonden aan de onderzoeksgroep 'Cognitive Psychology and Cognitive Science' van de universiteit van Exeter in Engeland.

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 midsagittal articulography
   Niels Schiller

3. Lexical access in the production of ellipsis and pronouns
   Bernadette 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 Austronesian language of Papua New Guinea
    Anna Margetts

13. From speech to words
    Arie van der Lugt