Processing the fine temporal structure of spoken words

ISBN: 978-90-76203-34-8

Cover illustration: Tilman Harpe
Cover design: GVO drukkers en vormgevers | Ponsen & Looijen, Ede
Printed and bound by: GVO drukkers en vormgevers | Ponsen & Looijen, Ede

Processing the fine temporal structure of spoken words

Een wetenschappelijke proeve

op het gebied van de Sociale Wetenschappen

Proefschrift

ter verkrijging van de graad van doctor

aan de Radboud Universiteit Nijmegen

op gezag van de rector magnificus prof. mr. S. C. J. J. Kortmann

volgens besluit van het college van decanen

in het openbaar te verdedigen

op dinsdag 29 juni 2010

om 15.30 uur precies

door

Eva Reinisch

geboren op 30 april 1982 te Graz, Oostenrijk

Promotores:	Prof. dr. A. Cutler
	Prof. dr. J. M. McQueen
Copromotor:	Dr. A. Jesse (MPI)

Manuscriptcommissie:	Prof. dr. H. J. Schriefers
	Prof. dr. V. J. J. P. van Heuven (Universiteit Leiden)
	Dr. L. C. Nygaard (Emory University)

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

Acknowledgments

Having finished the other parts of this book – THE book – there is one more thing I would like to say: Thank you so much everyone who made my life over the last few years such an exciting experience and who helped me along the way in putting together this volume. Thank you for the cheerful comments, advice, and discussions. Thank you for the suggestions and practical help with designing and setting up experiments, programming, finding bugs in my scripts, analyzing my data, translating things into Dutch, writing abstracts, articles, and much much more. The list goes on! Importantly, thank you for all the fun I had - at the institute and outside of work - and thank you for the encouragement you provided in times of need.

There are many people I am particularly grateful to for 'being there' to help and support me in all sorts of situations over these last years. Let me name just a few of you:

- my promotor Anne Cutler;
- my supervisors Alexandra Jesse and James McQueen;
- Holger Mitterer with his programming tricks and statistical advice;
- my roommates Susanne Brouwer (also paranimpf), Caroline Junge, and Attila Andics;
- other fellow PhD students: Matthias Sjerps (my second paranimpf), Miriam Ellert, Marijt Witteman, Begonia Diaz who visited the MPI several times;
- Rian Zondervan and the wider MPI administrative team;
- Ad Verbunt as contact person for technical help;
- the speakers in my experiments who read the endless lists of sentences I had to record: Marieke Pompe, and Marloes van der Goot;
- the L&C theater group;

Thank you to all members of the Comprehension Group for helping with bigger or smaller problems (like when I needed to know yet another thing about Dutch spelling or whether a certain word exists). I also greatly enjoyed our lunch conversations which provided welcome breaks between experiments, stats, and writing.

Thank you to Holger Mitterer, Marijt Witteman, Susanne Brouwer, and Mirjam Broersma for improving the Dutch summary of my thesis, and to Tilman Harpe for drawing the wonderful pictures for the cover of this book.

A big thanks also goes to Lynne Nygaard who kindly hosted me at Emory University for six months. Thanks to the students in the speech perception lab, and the members of the language group there. I really had a great time and I won't forget our discussions or the amounts of scrumptious cake we ate.

Vielen Dank auch an meine Eltern. Danke, dass ihr immer an mich glaubt und, was auch immer ich mir in den Kopf gesetzt habe, unterstützt. Danke an meinen Bruder und alle meine Freunde in verschiedenen Ecken Europas. Dziękuje bardzo! Danke für die vielen fröhlichen Chats, E-Mails, Karten und natürlich Besuche! Eure Geschichten, Diskussionen und Kommentare waren stets eine willkommene Abwechslung zum Alltag in Nijmegen.

Contents

Introduction

Chapter 1

Spoken language unfolds over time in a continuous stream of sounds. In contrast to written words, spoken words do not become available all at once, but gradually unfold with the speech signal. To recognize words, listeners have to rapidly process this incoming acoustic information. They continuously evaluate which word in the mental lexicon best matches the current acoustic input. The speech signal, however, unfolds at a variable pace. Utterances can be spoken faster or slower. Even though the length of all segments is affected by changes in rate, some segments are more affected than others (Crystal & House, 1982, 1988; Gay, 1978). Listeners have to take speaking rate into account when interpreting sounds (see Miller, 1981, 1987 for overviews). The English sound /b/, for example, has a shorter voice onset time (i.e., the time span between the release of the stop closure and the start of vocal-fold vibration) than the sound /p/. Nevertheless the sound /b/ can be perceived as /p/ if it is placed into a fast speaking rate context. Following a fast context, the voice onset time of the /b/ sounds relatively longer and thus leads to the interpretation of /p/. Duration is therefore interpreted in relation to its contextual influences. The aim of this thesis was to gain insights into how listeners use and evaluate temporal information during the processing of spoken words.

Speaking rate, however, is not the only temporal variation in the speech signal. The duration of sounds can vary for several reasons. First, sounds can be intrinsically long or short due to articulatory constraints. For example, the sound /m/ is longer than the sound /n/ because lip movements (as in /m/) are less flexible than tongue movements (as in /n/). Second, duration can cue sound identity, such as in phonemically long vs. short vowels (e.g., in Dutch 'taak' vs. 'tak' – "task" and "branch") or voiced vs. voiceless consonants (e.g., English 'bin' vs. 'pin'). Third, duration is part of a word's suprasegmental structure. The same sound can be spoken longer or shorter depending on its prosodic position in the word. The word-initial sound [ɔ] in the Dutch word 'octopus' ("octopus") is longer than the same sound in the initial position of 'oktober' ("October") because in 'octopus' it occurs in a stressed syllable,

but in 'oktober' the first syllable is unstressed (e.g., van Donselaar, Koster, & Cutler, 2005). Moreover, sounds at word boundaries tend to be longer if they are word initial than word final. For example, in the Dutch ambiguous sequence 'een(s) (s)peer' ("on(c)e (s)pear") the boundary sound [s] is more likely to belong to 'speer' the longer it is (Shatzman & McQueen, 2006). Fourth, the duration of sounds depends on their position in an utterance. Sounds or syllables in phrase-final position tend to be lengthened (e.g., Wightman, Shattuck-Hufnagel, Ostendorf, & Price, 1992). This shows that the temporal variation of speech sounds is not random, but rather determined by multiple influences.

Nevertheless, listeners interpret the duration of segments with regard to this variation seemingly effortlessly, even when several of these cues are combined. For example, the sound [ɔ] of the word 'octopus' is lengthened as it is word initial and stressed even though it is a phonemically short vowel. Still it can be recognized as the same vowel [ɔ] in the Dutch word 'radiator' where it is unstressed and not word-initial and thus shorter than in 'octopus'. Among all the possible types of durational variation that exist, this thesis focused on the perception of durational cues to a word's suprasegmental structure, specifically, cues to lexical stress and to word boundaries.

Lexical stress patterns are an interesting topic for the investigation of temporal processing. Unlike durational information about segment identity, lexical stress patterns are distributed over the syllables of a word. The Dutch words 'octopus' and 'oktober', for example, differ in the degree of stress not only on their initial syllables but also on their second syllables. That is, whereas the initial [ɔ] is stressed in 'octopus' but unstressed in 'oktober', the second vowel [o] is stressed in 'oktober' and unstressed in 'octopus'. This alternation of stressed and unstressed syllables in a word means that the processing of stress patterns must work over a larger temporal scale (i.e., over the syllables of a word) than when only segmental durations are at stake.

Dutch lexical stress provides a good window on the temporal processing of lexical stress because in Dutch it is possible to explore stress perception without interference from phonemic effects. In English, the vowels in unstressed syllables are mostly reduced and thus differ in segmental quality from stressed vowels. In contrast, Dutch lexical stress is mainly marked suprasegmentally. Dutch stressed vowels are longer, louder, have a higher pitch, and have more energy in high frequency bands than unstressed vowels. Duration is the most reliable cue to Dutch lexical stress (Nooteboom, 1972; Sluijter & van Heuven, 1995, 1996). Unlike pitch cues which are dependent on sentence intonation, durational cues are reliable under different sentence-focus conditions. Stressed and unstressed syllables can be

distinguished by their duration even when the sentence focus is shifted to an unstressed syllable (Sluijter & van Heuven, 1995). Given the importance of duration as a cue to lexical stress in Dutch it makes it possible to investigate temporal processing over a larger time scale while controlling for segmental properties.

The investigation of the temporal processing of lexical stress patterns is interesting for another reason. As mentioned above, multisyllabic words have alternating patterns of stressed and unstressed syllables. The different degrees of stress, however, vary for different words. As mentioned above, 'OCtopus' and 'okTOber' (capitals indicate stress) have primary stress on the first vs. second syllable (1-2 contrast) and their initial syllables differ in primary stress vs. no stress. In contrast, word pairs of the type 'DIameter' and 'diaMANT' ("diameter" – "diamond") have primary stress on the first vs. the third syllable (1-3 contrast) and their initial syllables differ in primary vs. secondary stress. Dutch words with primary stress on the third syllable (i.e., 'diaMANT') have secondary stress on their initial syllables. Primary stressed syllables are longer than secondary stressed syllables and secondary stressed syllables are longer than unstressed syllables (Rietveld, Kerkhoff, & Gussenhoven, 2004; Slootweg, 1988). Although listeners can distinguish between these different degrees of stress (Mattys, 2000) it is possible that word pairs which differ in primary vs. secondary stress on their initial syllables are harder to distinguish than words from the other stress contrast. Words from the 1-3 stress contrast could thus suffer more from lexical competition and be recognized later than words from the 1-2 stress contrast.

Previous studies suggest that Dutch listeners use lexical stress information to recognize words (e.g., Cooper, Cutler, & Wales, 2002; Cutler & van Donselaar, 2001; van Donselaar et al., 2005; van Heuven, 1988; van Leyden & van Heuven, 1996). The use of lexical stress information can lead to an earlier recognition of words than when stress is not taken into account (Cutler & Pasveer, 2006; van Heuven & Hagman, 1988). Statistical computations on the Dutch lexicon suggested that taking stress information into account reduces the number of embedded words from an average of 1.52 to 0.74 words (Cutler & Pasveer, 2006). Moreover, with stress information considered, Dutch words become unambiguous after hearing an average of only 67% of their phonemes compared to 80% if stress information is ignored (van Heuven & Hagman, 1988). This statistical advantage of using stress information in word recognition is also reflected in Dutch listeners' behavior (Cooper et al., 2002; Cutler & van Donselaar, 2001; van Donselaar et al., 2005; van Heuven, 1988; van Leyden & van Heuven, 1996). When guessing a word from its first syllable, listeners reported the correct degree of stress for that syllable in 80% of the cases (van Leyden & van Heuven, 1996). Similarly, in a fragment priming task, listeners responded

faster to a printed target 'octopus' if they heard the segmentally and suprasegmentally matching fragment prime 'OCto-' than an unrelated fragment 'eufo-' (van Donselaar et al., 2005). This facilitation was even found with one-syllable primes (i.e., 'OC-'). In contrast, a fragment that matched the target segmentally but mismatched the target in stress (i.e., 'okTO-' as in 'okTOber') inhibited listeners' responses to 'octopus'. Inhibition was found with two-syllable primes ('-okTO') but not with one-syllable primes (i.e., 'ok-'). These prior studies suggest that stress information is used to recognize words and that the amount of stress information available (i.e., one vs. two syllable primes) plays a role.

Building on this prior evidence that listeners use suprasegmental cues to recognize words, the experiment reported in Chapter 2 focused on the time course of processing lexical stress information as the signal unfolds. Prior studies give a first indication about the amount of information needed to use a stress pattern for word recognition. In a gating task, as used by van Leyden and van Heuven (1996), listeners hear a fragment of a word and then indicate the word they heard. Gating thereby systematically varies the size of the presented fragment. As mentioned above, after approximately one syllable the stress pattern can mostly be recognized. For the question about the processing of words in real time, however, gating is less suitable. Gating may reflect post-perceptual processing rather than the word recognition process itself. Priming — the task used by van Donselaar et al. (2005) — is likely to tap into earlier phases of word recognition. The fact that one-syllable primes facilitate recognition of the target but do not inhibit segmentally competing words suggests that stress information on the initial two syllables is critical to recognize words. A more fine-grained time course analysis of when stress information is used, however, is not feasible with a priming paradigm.

To address precisely when stress information is used during word recognition, the experiment reported in Chapter 2 used printed-word eye tracking. In this task, participants listen to a sentence while seeing four printed words on a screen. Eye tracking makes it possible to monitor lexical competition over time. It exploits the fact that listeners spontaneously fixate visual referents to acoustic input (Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Visual referents can be objects, pictures, or printed words (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Importantly, eye movements to these referents are closely time-locked to the speech signal and thus reveal the time course of lexical competition. When listeners, for example, hear the Dutch word 'alibi' ("alibi"), they initially also look at the word 'alias' ("alias") since both words start with the same segments /ali/. But once disambiguating information comes available - in the example here the segment /b/ - fixations on the competitor 'alias' decrease and the target 'alibi' is recognized.

Eye-tracking studies suggest that all words that match the current acoustic input compete for recognition. Previous eye-tracking studies demonstrated how segmental properties of this input affect the competition process over time (e.g., Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Tanenhaus, et al., 1995). The experiment described in Chapter 2 asks whether and how suprasegmental information about lexical stress modulates competition. The words 'alibi' and 'alias' overlap segmentally on their initial two syllables, as do the words 'alibi' and 'alinea' ("paragraph"). But although the words 'alibi' and 'alinea' overlap segmentally on their initial two syllables, they additionally differ in lexical stress placement. 'Alibi' has primary stress on the first syllable, 'aLInea' has primary stress on the second syllable (1-2 contrast). So segmental information about the /b/ at the onset of the third syllable is not the earliest acoustic information that disambiguates these words. The question was whether listeners use such stress differences to distinguish between words as soon as stress cues come available and before segmental information distinguishes the words. If listeners optimally use suprasegmental information as soon as it comes available, they should use lexical stress information to disambiguate words even if that information comes available earlier than disambiguating segmental information. As noted above, the amount of competition may depend on the stress patterns of the words (i.e., 1-2 vs. 1-3 contrast). If the initial (two) syllables of the words in a pair are acoustically similar - as it is the case for the 1-3 contrast - the words may suffer from stronger competition and be recognized later than words with larger stress differences (stressed vs. unstressed syllables in the 1-2 contrast). Eye tracking reveals the time course of this competition process. Chapter 2 thus addressed yet another aspect of the temporal processing of speech: when does incoming acoustic evidence modulate lexical competition?

Chapters 3 and 4 also addressed the perception of temporal structure in Dutch lexical stress patterns. As mentioned above, the interpretation of temporal information crucially depends on the speaking rate of the utterance. Chapters 3 and 4 therefore explored how listeners interpret lexical stress patterns relative to the speaking rate of a preceding context. Most prior evidence showing that listeners use speaking rate information in word recognition comes from studies on phoneme perception (e.g., Allen & Miller, 2001; Miller, 1981; 1987; Miller & Dexter, 1988; Miller & Liberman, 1979; Miller & Wayland, 1993; Wayland, Miller, & Volaitis, 1994). These studies mainly focused on speaking rate effects on category structure. For example, changes in speaking rate context not only affect the location of the categories' best exemplars but also the width of the best exemplar range (Wayland, et al., 1994). To address this type of question, these studies used phoneme categorization and

goodness judgment tasks. By investigating speaking rate effects on phoneme perception, they focused on local durational cues.

The research reported in Chapters 3 and 4 extended these findings. By investigating speaking rate effects on lexical stress, rate effects on temporally distributed cues were explored. In addition, Chapter 3 investigated the time course of processing rate information during stress-modulated lexical competition. The joint investigation of suprasegmental cues and the influence of speaking rate should therefore further contribute to a closer understanding of the inner workings of the word recognition process. Although it has repeatedly been shown that listeners optimally evaluate upcoming acoustic information at any point in time (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Norris & McQueen, 2008; Tanenhaus, et al., 1995) little is known about when temporal context information is used during the word recognition process.

Previous studies on phoneme perception suggested that speaking rate effects occur at an early, prelexical phase of processing. Some evidence for the earliness of rate effects comes from studies that demonstrated the use of speaking rate information from speakers other than the target speaker (Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000). Rate information is used even from speakers at a different spatial location than the target speaker's location (Newman & Sawusch, 2009). Speaking rate information thus cannot be ignored and is taken into account before other early processes such as perceptual grouping or stream segregation occur. It has to be noted, however, that these studies once again focused on temporal processing of segmental information.

The experiments reported in Chapter 3 therefore investigated whether evidence for such an early use of speaking rate can also be found in temporal processing of suprasegmental information. While Chapter 2 addressed whether lexical stress information is used in word recognition as the speech signal unfolds, Chapter 3 went on to ask, with the same materials and task, whether speaking rate influences the use of stress information immediately during lexical competition. Although duration is the most reliable cue to lexical stress in Dutch, a slow speaking rate context should make a word's initial syllable sound relatively shorter and thus less stressed than following a fast rate context. If listeners perceive the degree of stress on the initial syllable of a word relative to the rate of the preceding context, non-initially stressed words (e.g., 'okTOber' and 'diaMANT') should compete more strongly for recognition when the initially stressed target ('OCtopus' and

'DIameter') is preceded by a slow than a fast context. The opposite was expected for non-initially stressed targets.

These questions were addressed in two experiments, in which also the ambiguity of local stress information in the target words was manipulated. In the first experiment, all cues to stress (i.e., duration, pitch, amplitude) were present and only the preceding carrier sentence was rate-manipulated. In a second experiment, stress cues other than duration were neutralized. This made it possible to explore speaking rate effects on durational cues that were salient (i.e., when duration is the only remaining stress cue, as in the second experiment) vs. less salient (i.e., when other stress cues are present, as in the first experiment). When duration is the only cue to stress, speaking rate effects should be stronger than when listeners have other stress cues to rely on that are not interpreted relative to rate.

Chapter 4 also addressed the use of speaking rate information in the perception of stress, but with another task, a fragment categorization task. One word pair of the 1-2 stress contrast ('Alibi' – 'aLInea'; "alibi" – "paragraph") and one word pair of the 1-3 stress contrast ('CAvia' – 'kaviAAR', "guinea pig" – "caviar") were chosen from the set of pairs used in the eye-tracking experiments. Listeners had to indicate from which word of a pair the segmentally overlapping initial two syllables had been taken. The goal of this series of experiments was to address the interaction of speaking rate effects and the presence of durational cues on initial and non-initial syllables of the words. In contrast to Chapter 3, where the time course of processing was central, here the focus was on effects of speaking rate on the different syllables of the words. So even though the categorization task may reflect post-perceptual processing rather than word processing over time, it was suitable for addressing the present questions about rate effects on different degrees of stress in different stress patterns.

In the experiments reported in Chapter 4 not only the speaking-rate context but also the durational cues on the initial and second syllables of the word pairs were manipulated. One of the fragments' syllables was manipulated along a duration continuum while the other syllable was set to a perceptually ambiguous duration. This allowed for a comparison of rate effects when the durations of the first and second syllables of the fragments were themselves informative about stress (i.e., manipulated along a duration continuum where only a few steps were ambiguous) or were not informative about stress (i.e., the syllables were consistently set to a perceptually ambiguous duration). It was predicted that listeners' categorizations reflect the use of a combination of rate-modulated stress cues on the initial

syllables and the durational information of whatever syllable was manipulated along a continuum. In that way even the categorization task could give some insight into the processing of rate information over time. If the uptake of acoustic information in word recognition is incrementally optimal (Norris & McQueen, 2008), rate effects on a word's initial syllable should be stronger than on non-initial syllables.

The experiments reported in Chapters 3 and 4 therefore addressed the influence of speaking rate context on stress perception in complementary ways. By using the same task and materials as Chapter 2, the experiments in Chapter 3 focused on the immediate effect of rate as a word unfolds over time. In Chapter 4, in contrast, a categorization task was employed to systematically explore conditions under which speaking rate affects the interpretation of the lexical stress carried on different syllables of a word. The two different methods should provide a comprehensive picture of the processing of temporal information over time. A comparison of online results obtained with eye tracking to categorization results where listeners had time to process the stimuli before they initiated a response should further capture the limits of word recognition.

Chapter 5 extended the investigation of rate effects in online and offline processing to suprasegmental durational cues to word boundaries. There are no reliable breaks in the unfolding speech stream that indicate when a word ends and hence where access to a new word in the mental lexicon can start. Rather, listeners have to use a variety of acoustic cues to segment speech (e.g., Cho, McQueen, & Cox, 2007; Mattys, White, & Melhorn, 2005; Nakatani & Dukes, 1977). As well as being a cue to lexical stress, duration is also an important cue to word boundaries (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, Liberman, Eccardt, & Pesetsky, 1978; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Burani, & Scott, 1995). Importantly, duration as a cue to word boundaries is used as the acoustic signal unfolds to modulate lexical competition (Shatzman & McQueen, 2006). When presented with sentences that contain ambiguous word sequences of the type described above (e.g., 'wel eens (s)peer', "once spear/ pear") and visual target-competitor pairs such as 'peer' and 'speen' ("pear", "pacifier") listeners consider 'speen' as a possible target longer the longer the boundary sound [s] is (Shatzman & McQueen, 2006).

Building on this prior evidence the first question asked in Chapter 5 was whether duration as a cue to word boundaries is also perceived relative to speaking rate. Repp et al. (1978) suggested that the perception of a word sequence such as 'grey chip' vs. 'great ship' depends on the durations of the segments surrounding the possible word boundaries.

Therefore a longer speaking rate context should shift the perception of word boundaries as well. In the example above ('wel eens (s)peer') listeners should longer consider the [s]-initial word 'speer' as a possible target if the preceding sentence context was spoken fast than when it was slow. By exploiting an eye-tracking paradigm to capture the time course of rate-induced lexical competition it was further possible to compare the processing of rate information over time to rate effects in offline processing such as these mentioned in categorization tasks. The time course of rate effects was expected to be similar to effects of durational cues on the boundary sounds themselves. These were the questions of Experiments 1 and 2 of Chapter 5.

A second series of experiments in Chapter 5 addressed the role of the amount and location of speaking rate information in segmenting speech. During word recognition listeners continuously update the current acoustic information used in lexical access. Therefore they are likely to update the speaking rate information they use as well. The question then arises as to how much speaking rate context listeners use to evaluate the current acoustic information. Moreover, where does this context have to be in order to be used? Speaking rate information could be taken from only the context that is proximal to the cues to be evaluated but it could also be calculated from a more distal context, for example, when the proximal context does not allow for the disambiguation of certain sounds.

Some previous studies suggested that rate information does not have to be adjacent to a target sound in order to affect its perception (Port, 1979; Gordon, 1988; Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). The word-medial consonant on a 'rabit'- 'rapid' continuum, for example, is interpreted relative to rate context even if the initial syllable had been spoken at a normal, constant, rate (Gordon, 1988). Effects of adjacent and non-adjacent rate information have been suggested to occur within a time window of approximately 300 ms preceding and following the target (Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). Other studies, however, found that under certain conditions speaking rate information from a more distal context than 300 ms can have an effect as well (Kidd, 1989; Summerfield, 1981; Wayland, et al., 1994). Kidd (1989), for example, found that the rate of distal stressed syllables in a sentence has a greater influence on phoneme perception than the rate of proximal but unstressed syllables. Wayland et al. (1994) further established that distal context shifts the perceived locations of the phoneme's best exemplar range but does not affect the width of this range as proximal context does. These studies thus established some conditions under which distal context can affect the perception of phoneme categories.

Chapter 5 attempted a more systematic investigation of the roles of proximal and distal rate context in word segmentation. As noted above, all previous studies on the use of speaking rate information used tasks in which listeners had time to process the context and targets before they initiated a response. It appears plausible, however, that listeners can use a longer or more distal speaking rate context when they have time to process the stimulus post-perceptually as compared to the continuous information update during word recognition. Since listeners continuously update rate information for the evaluation of upcoming cues they could give more weight to proximal context during word recognition. Chapter 5 therefore employed eye tracking as well as categorization tasks to explore the role of proximal and distal rate contexts. The experiments asked about possible differences in the use of rate information when listeners have little (eye tracking) or more (categorization) time available to process the speaking rate context.

To assess to what degree the use of distal context may depend on processing time, the carrier sentences were divided into a proximal context ('wel eens' in 'Ze heeft wel eens (s)peer gezegd' – "She once said (s)pear") and a distal context (i.e. all context preceding 'wel eens'). The proximal context was of maximally 250 ms duration at a slow rate. Basic effects of distal context should thus be observed even if the previously suggested 300 ms were the limit of rate context that could be taken into account. The question was, whether on top of these expected effects, further manipulations would modulate the results. When the proximal context had the opposite speaking rate than the distal context an attenuation of the effect of proximal context was expected (Experiment 3). Experiments 4 and 5 further addressed whether, in addition to the location of the rate context, its amount plays a role. The question was whether a distal context could compensate for its distal location by increased length, and whether a long distal context leads to stronger rate effects than a short distal context. If listeners can use more rate information when more processing time is available (i.e., in categorization tasks) the length of the rate context could matter and even compensate for a distal location. During the word recognition process, however, proximal context should count more. The manipulation of proximal and distal rate information should thus inform us about the processing of speaking rate information in online (eye tracking) vs. offline (categorization) word recognition.

In summary, the goal of this thesis was to gain insight into how listeners use temporal information during spoken word recognition. The focus was on the processing of a words' suprasegmental structure, in the form of durational cues to lexical stress and to word boundary location. Since durational information, however, is not perceived in isolation, the perception of lexical stress and word boundaries relative to the speaking rate of the

preceding context was also explored. The use of online (eye tracking) and offline (categorization) tasks addressed the role of temporal information during word recognition vs. at a late, post-lexical phase of word processing. The comparison of speaking rate effects from proximal and distal contexts in online and offline processing should give further insight into the mechanisms of rate evaluation in different processing situations. This thesis thus explored the processing of the fine-temporal structure of spoken words.

Early use of phonetic information in spoken word recognition:

Lexical stress drives eye movements immediately

Chapter 2

Reinisch, E., Jesse, A., & McQueen, J. M. (2010), Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately. *The Quarterly Journal of Experimental Psychology, 63*, 772-783.

Abstract

For optimal word recognition listeners should use all relevant acoustic information as soon as it comes available. Using printed-word eye-tracking we investigated when during word processing Dutch listeners use suprasegmental lexical stress information to recognize words. Fixations on targets such as 'OCtopus' (capitals indicate stress) were more frequent than fixations on segmentally overlapping but differently stressed competitors ('okTOber') before segmental information could disambiguate the words. Furthermore, prior to segmental disambiguation, initially stressed words were stronger lexical competitors than non-initially stressed words. Listeners recognize words by immediately using all relevant information in the speech signal.

Introduction

To comprehend spoken language, listeners have to determine which words a speaker said. This is not a trivial task, since it involves establishing which of the approximately 50,000 entries in the mental lexicon best matches the information provided by the speech signal. All words that temporarily match the signal compete for recognition. The word that receives the most support is most likely to be recognized. Word recognition is further complicated by the fact that information about what was said is not available at once but rather is provided over time. To deal with these temporal demands efficiently, listeners use incoming information about the segments of words to select among competing lexical hypotheses as it comes available (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Norris & McQueen, 2008; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

The speech signal, however, not only provides segmental information but also suprasegmental information, such as duration, pitch, and amplitude. For example, suprasegmental stress patterns differentiate the meanings of the English words "(a) FORbear" and "(to) forBEAR" (capital letters indicate stress). But listeners do not always rely on suprasegmental information to resolve lexical competition (Cutler, 1986). This may be because word pairs like "forbear" are rare. In the majority of cases differences in segmental information are sufficient to recognize words. In addition, suprasegmental information tends to come available later than segmental information (Cutler & Chen, 1997). For example, a larger part of the signal is needed to perceive pitch movement than to perceive vowel quality, since pitch changes can be perceived only over time. However, if word recognition is based on the uptake of all acoustic information as soon as it comes available, listeners should use suprasegmental information when it can distinguish words earlier than segmental information. We asked here whether this is indeed the case.

We examined when, during word recognition, Dutch listeners use suprasegmental stress information to resolve lexical competition. Dutch lexical stress provides a good test case for two reasons. First, unlike in English, where most unstressed vowels are reduced, lexical stress in Dutch is mainly marked suprasegmentally. That is, whereas in English the first vowels of 'octopus' and 'October', for example, are qualitatively different (stressed [ɔ] in 'octopus', unstressed [ə] in 'October'), the first vowels in the respective Dutch words 'octopus' and 'oktober', and indeed the next three sounds, are segmentally the same ([ɔkto]). We could therefore ask if Dutch listeners can use the stress differences between, for example, 'OCtopus' and 'okTOber' before they hear the segmental difference (/p/ vs. /b/). Second, although few Dutch words are contrasted by stress only, Dutch listeners could benefit from

early uptake of stress information (Cutler & Pasveer, 2006). Whereas Dutch words contain on average 1.52 embedded words when stress is ignored, this reduces to 0.74 embedded words when stress is taken into account. Furthermore, stress information in the lexicon shifts the average segmental uniqueness point considerably nearer to the beginnings of words (van Heuven & Hagman, 1988). With stress information considered, words can be distinguished on average after only 67% of their phonemes, instead of after 80% without stress taken into account. The use of stress information in Dutch word recognition would therefore be beneficial.

Dutch listeners indeed appear to use lexical stress information. In gating studies, listeners are presented with subsequently longer parts of a word for recognition. As little as hearing the first syllable of a word is sufficient to recognize the stress status of the first syllable in 80% of the cases (van Leyden & van Heuven, 1996). Cross-modal repetition priming experiments with word fragments have shown that listeners' responses to targets are influenced by the stress pattern of the primes (Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005). When the stress pattern of the heard fragment (e.g., 'OCto') matched the pattern of the printed target ('octopus'), listeners' decisions were faster relative to an unrelated control condition. This was the case even when the prime consisted of only one syllable. An inhibitory effect on the recognition of targets that mismatched their prime's stress pattern, however, was found only with two-syllable primes (i.e., 'okTO-', 'octopus').

These prior studies, however, leave open the question when stress information is used during word recognition. In gating studies, listeners have time to make their guess about the word they hear. Stress information could therefore be considered only relatively late, that is, during post-perceptual decision-making. In the priming paradigm, effects of stress measured on the visual lexical decisions are likely to be based on the perceptual processing of the prime rather than on post-perceptual processes. But although the use of fragment primes of varying lengths (i.e., one vs. two syllable primes) gives an indication of the amount of information needed to facilitate the recognition of words, it does not allow continuous evaluation of the time-course of the competition process. In particular, since the visual decisions are made after the acoustic offsets of the primes, it remains unclear when exactly stress information is used during the processing of the primes.

In the present study, therefore, we looked at the earliest moments of spoken-word recognition using the printed-word eye-tracking paradigm (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Listeners spontaneously fixate visual referents to auditory input

(Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974). Critically, the timing of eye-movements is closely linked to the timing of the acoustic signal and thus reflects the degree of support for lexical candidates over time. For example, when listeners hear the beginning of the word 'beaker' they initially also look at a picture of a beetle because both words begin with the same segments. As distinct segmental information (i.e., [k]) comes available, listeners look at the target picture beaker more frequently than at the beetle (Allopenna et al., 1998). If suprasegmental stress information is also used as it comes available, listeners should differentiate segmentally overlapping words on the basis of a different stress pattern before segmental information can distinguish between the words. We thus asked whether fixations on an initially stressed target such as 'OCtopus' are more frequent than fixations on its segmentally overlapping but non-initially stressed competitor 'okTOber' before the words are segmentally distinct.

The investigation also had a second purpose. In gating experiments listeners give more initial stress responses than non-initial stress responses (van Leyden & van Heuven, 1996). This response asymmetry could be due to the distribution of stress locations in the Dutch lexicon. More Dutch words have word-initial stress than non-initial stress (van Heuven & Hagman, 1988). The asymmetry could thus arise from a decision-level bias that considers this prior distribution in the language. Previous priming experiments have not addressed this issue. In a series of stress categorization experiments, however, van Heuven and Menert (1996) asked whether acoustic context characteristics can influence this response asymmetry. They suggested that the initial stress bias in many experiments may be due to the presentation of words in isolation rather than in sentence contexts. For words presented in isolation, the lack of preceding acoustic context induces a perceived pitch rise on the initial syllable which is consequently interpreted as initial stress. This suggests that the initial stress bias is not solely due to statistical knowledge, but is also at least partially driven by acoustic information, at least by information provided by the context. The present study tested whether acoustic information on the words themselves contributes to the response asymmetry and hence whether this asymmetry emerges from early perceptual processes. In particular, since a syllable is perceived as stressed if it is perceived as prominent relative to its context (van Heuven & Menert, 1996), the presence of stress cues could be more informative than their absence. If stress cues are clearly present in the speech signal, initially stressed words would receive more support than non-initially stressed words. The absence of cues, however, results in greater uncertainty since stress cues could have been reduced by the speaker or missed by the listener. Examining the time-course of word recognition in eye-tracking could therefore reveal the locus of the previously observed response asymmetry. If

this asymmetry is signal driven, stress location on the target word should modulate the time-course of competition for recognition.

Method

Participants

Twenty-four Dutch native speakers with no reported hearing problems and normal or corrected-to-normal vision were paid for taking part.

Stimuli

Twenty-four three- or four- syllable stress pairs served as targets (see Appendix). The words of a stress pair were segmentally identical for at least their first two syllables but differed in the location of primary stress. Seven pairs had stress on the first or the second syllable (e.g., 'OCtopus', 'okTOber'; 1-2 stress contrast). Seventeen pairs had primary stress on the first or the third syllable (e.g., 'CENtimeter', 'sentiMENT'; 1-3 stress contrast). Dutch words with primary stress on the third syllable have secondary stress on the first syllable. Six pairs from each contrast set could be distinguished segmentally at the end of the second syllable, the others within the first two phonemes of the third syllable. Eight additional word pairs were chosen with similar criteria to serve as fillers and six more for practice trials. Each word pair was presented on a computer screen together with a phonetically and orthographically dissimilar distractor word pair. Words in the distractor pairs were segmentally overlapping with each other in their first two syllables but did not necessarily differ in stress placement (e.g., 'diaLECT' and 'diaLOOG'). Words from the stress and distractor pairs were semantically unrelated and matched for log-transformed frequency (CELEX; Baayen, Piepenbrock, & Gulikers, 1995) within ($t(23) = -0.26$, $p = .80$) and across pairs ($t(94) = -0.83$, $p = .41$).

Recording

A female Dutch native speaker was recorded in a sound-attenuated room. Target words were uttered at the end of the sentence 'Klik nog een keer op het woord' ("Click once more on the word") with sentence accent on the target. Average sentence duration without targets was 1200 ms.

Procedure

Participants were seated approximately 60 cm in front of a 32.5 by 24 cm screen. First, they were familiarized with the stimuli. All words were presented in lower-case one after the other in the middle of the screen and participants were required to read them aloud. No feedback was given. The eye-tracking experiment followed immediately. Eye-movements were recorded with a head-mounted SMI Eyelink II System at a sampling rate of 250 Hz.

During the main experiment participants saw 32 displays with four printed words repeated four times, each shown once in each of four blocks. Across blocks an answer to each word of a stress pair was obtained from every participant. In the first block, the targets were words from each of the stress pairs (experimental and filler pairs) chosen at random such that half had initial stress. In subsequent blocks, a target could either be the same word as before, its segmentally overlapping competitor, or a word from the other pair. It was thus unpredictable which word would be the next target when a display was repeated. The number of words from each stress contrast and stress location was the same across blocks. Order of blocks was counterbalanced across participants. Order of trials within a block was randomized separately for each participant. There were no breaks between blocks. The first block was preceded by twelve practice trials that consisted of six displays, each presented twice.

On every trial participants saw a fixation cross for 500 ms centered on the screen. 200 ms later four printed words appeared for 2400 ms. All words were presented in monospaced lower-case Lucida Sans Typewriter font, size 20, centered in the four quadrants of the screen. The average-length word covered approximately 3.18 degrees of visual angle. Auditory instructions (i.e., carrier sentences plus targets) were played over headphones at a comfortable listening level. The acoustic onset of the sentence was timed such that the onset of the target word was 1200 ms after the printed words appeared on the screen. The participants' task was to click with the mouse on the target word. The experiment contained 140 trials and lasted approximately 15 minutes. Every fifth trial a drift correction was carried out to adjust for possible head movements.

Results

Only trials in which participants clicked on the correct word were analyzed. Only eight trials (0.7%) had to be excluded for this reason. If a target was repeated during the experiment, only data from the first presentation were used. Fixations on a word were counted as such if they fell within a predefined square of 6.3 cm side length, centered around the middle of each word.

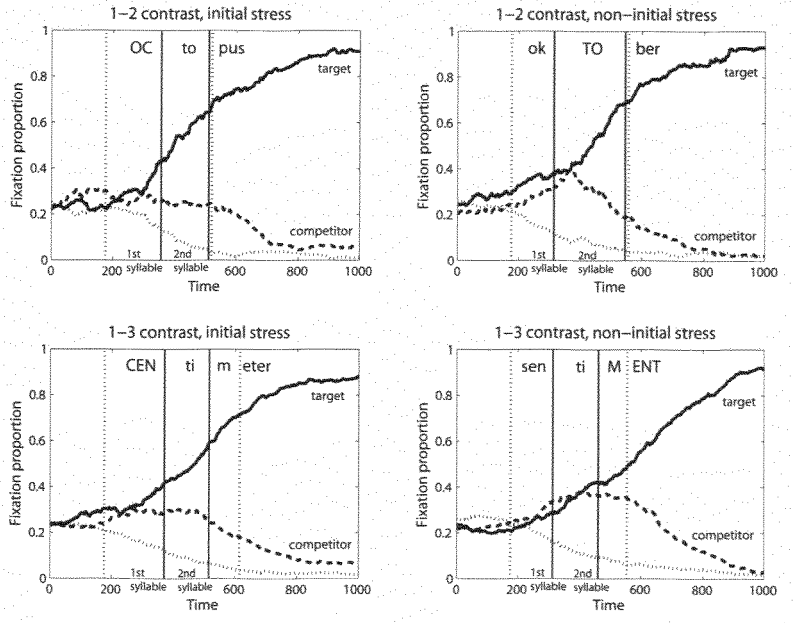

Figure 1. *Fixation proportions over time to target (solid line), competitor (dashed line), and averaged distractors (dotted line) from acoustic target onset. Vertical dotted lines indicate the critical time window (see text for details). Solid vertical lines show normalized and time-shifted syllable offsets (i.e., for each item the same number of events within one syllable was taken and plotted aligned with the timeline of the average syllable durations, and these measures were shifted by 180 ms, as for the critical time window).*

Figure 1 shows fixation proportions on target, competitor, and the average of the two distractors over time for each of the four conditions defined by stress contrast and stress location. The dashed vertical lines mark the average critical time window in which stress but not segmental information could distinguish the words of a stress pair. It encompassed the time from target onset to the point at which the target segmentally diverged from its competitor shifted by an estimate of the time needed to program and launch a saccade (see, e.g., Matin, Shao, & Boff, 1993). This estimate was defined as the amount of time from word onset required for fixations on the segmentally mismatching distractors to become less frequent than fixations on target or competitor. T-tests on fixation proportions in consecutive 4 ms time windows from target onset (i.e., on every time sample provided by the eye-tracker) revealed significantly more fixations on target and competitor than distractors from 180 ms after target onset onwards (see Figure 2). The critical time window was therefore defined as extending from 180 ms after target onset to each word's segmental target-competitor divergence point plus 180 ms.

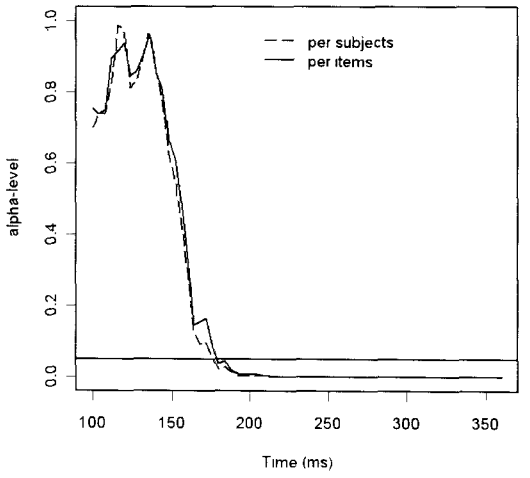

Figure 2. *Point-to-point analysis in consecutive time windows of 4 ms. Plot of p-values (i.e., probability that there is no difference in fixation proportions between distractors and targets or competitors) over time from target onset. The horizontal line indicates the significance level of* $\alpha = .05$.

Critical time window

Analyses of variance by participants (F1) and by items (F2) were run with stress contrast (1-2 and 1-3 contrast) and stress location (primary stress on the first syllable or not) as within-participant and between-item factors. Analyses were run separately for fixation proportions on targets and, as a measure of competition, for the difference between fixations on the competitor and the average of the two distractors. Since the factor 'block' neither approached significance nor interacted with any of the other factors (all $p > .47$) it was dropped from all analyses. Results are summarized in Table 1. For fixations on the targets no main effects were found in the critical time window. However, there was an interaction between stress contrast and stress location. Follow-up comparisons showed that non-initially stressed targets from the 1-2 stress contrast received more fixations than non-initially stressed targets from the 1-3 contrast ($F1(1,23) = 11.08$, $p < .005$; $F2(1,22) = 7.71$, $p < .05$) but there was no difference between the two types of targets with initial stress (F1 and $F2 < 1$). The analysis of competitor-distractor differences showed that competition was affected by stress location. Words with initial-syllable stress were stronger competitors than words with non-initial stress.

By-syllable analyses

The data were then divided into time windows that corresponded to the first and second syllables of the targets (see Table 1). In the analyses of fixations on the target the interaction of contrast and location was found only on the first syllable. Stress contrast again affected the looks to non-initially stressed targets ($F1(1,23) = 7.42$, $p < .05$; $F2(1,22) = 10.54$, $p < .005$) but not to initially stressed targets (F1 and $F2 < 1$). On the second syllable an effect of stress contrast was found. Targets from the 1-2 contrast were fixated more frequently than words from the 1-3 contrast. In the analyses of competition, the stress location effect established in the overall analysis (i.e., words with initial-syllable stress were stronger competitors than words with non-initial stress) was found on the second syllable only. On the first syllable no effect of contrast or location was apparent.

		target fixations					
		contrast		location		contrast*location	
		F	p	F	p	F	p
critical time window	F1(23)	.84	= .36	.003	= .96	8.54	< .01
	F2(44)	.68	= .41	.008	= .93	5.19	< .05
first syllable	F1(23)	1.53	= .23	.003	= .96	11.76	< .005
	F2(44)	2.11	= .15	.006	= .94	6.59	< .05
second syllable	F1(23)	5.69	< .05	.34	= .57	2.44	= .13
	F2(44)	6.55	< .05	.35	= .56	2.17	= .15
		fixations on competitor minus distractors					
critical time window	F1(23)	.50	= .49	5.91	< .05	1.73	= .20
	F2(44)	.25	= .62	3.40	= .072	1.23	= .27
first syllable	F1(23)	.005	= .95	.52	= .48	2.20	= .15
	F2(44)	.001	= .98	.37	= .55	1.09	= .30
second syllable	F1(23)	.29	= .6	6.98	< .05	.91	= .35
	F2(44)	.13	= .72	3.80	= .058	.87	= .36

Table 1. *Effects of stress contrast and stress location and the interaction of contrast by location on fixations on the target and, as a measure of competition, the difference in fixations on the competitor and the averaged distractors.*

Target-competitor disambiguation

The critical hypothesis was whether there was a preference for fixating the target over the competitor before the word pairs became segmentally distinct. An analysis of the ratio of fixations on the target to the sum of fixations on target and competitor was conducted in the critical time window (see Table 2). Looks to the target were more frequent than looks to the competitor before disambiguating segmental information became available. Only targets with primary stress on the third syllable were on average not fixated more frequently than their competitors. A point-by-point analysis was carried out to establish when fixations on the target became more frequent than fixations on the competitor.

	critical time-window							
	1-2 stress contrast				1-3 stress contrast			
	initial stress		non-initial stress		initial stress		non-initial stress	
	t1(23)	t2(6)	t1(23)	t2(6)	t1(23)	t2(16)	t1(23)	t2(16)
t	2.03	2.86	2.96	2.49	3.54	3.23	1.65	1.29
p	= .054	< .05	= .007	< .05	< .005	< .005	= .11	= .22

Table 2. *Preference of fixations on the target over the competitor per condition in the critical time window from 180 ms after target onset to the segmental target-competitor divergence point plus 180 ms.*

	1-2 stress contrast				1-3 stress contrast			
	initial stress		non-initial stress		initial stress		non-initial stress	
onset of last shared segment	68%		69%		79%		65%	
	t1(23)	t2(6)	t1(23)	t2(6)	t1(23)	t2(16)	t1(23)	t2(16)
target - competitor disambiguation	59%	58%	71%	68%	57%	54%	70%	86%

Table 3. *Time of the onset of the last shared segment of target and competitor and the point in time at which fixations on the target became more frequent than fixations on the competitor ($p < .05$, by participants, t1, and items, t2) and remained so for at least 20 time slices. Time is given in percent of the critical time window per condition.*

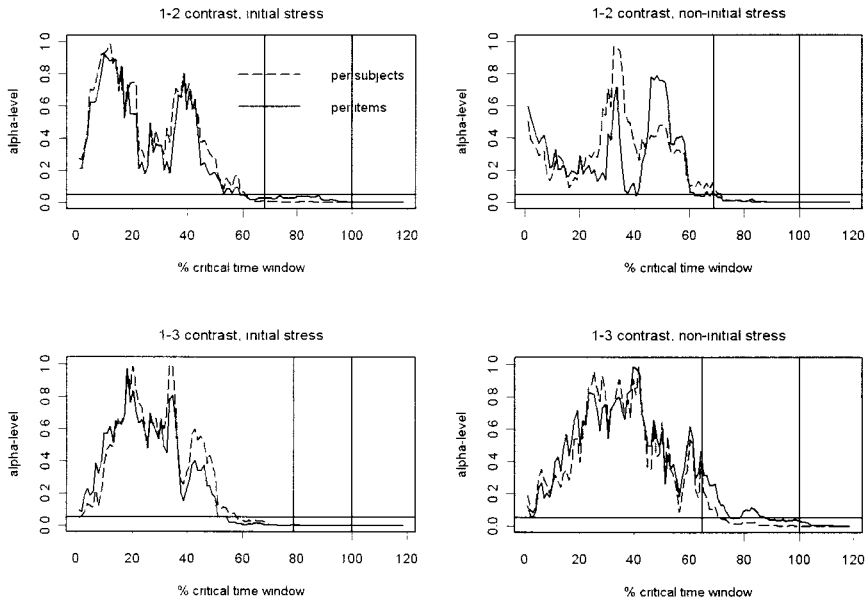

Figure 3. *Point-to-point analysis of target - competitor divergence per condition. Plot of p-values over time in percent of the critical time window (end of critical time window = 100%). The horizontal line indicates the significance level of* $\alpha = .05$*, the left vertical line indicates the average onset of the last shared segment per condition, the right vertical line indicates the segmental divergence point of target and competitor (i.e., 100%).*

To normalize the time to the segmental target-competitor divergence point across items, 98 equally-spaced time windows were created in the critical time window for each item (98 = duration of average critical time window/4). In all four conditions, targets were fixated more frequently than competitors before the words became segmentally distinct (see Figure 3 and Table 3). Moreover, fixations on the target were more frequent than fixations on the competitor before the onset of the last shared segment of the stress pair for targets with initial stress, and at the onset of this segment for targets with primary stress on the second syllable.

Acoustic measures

To explore what information listeners used to determine the stress pattern of the targets, acoustic measurements of duration (ms), mean pitch (Hz), spectral tilt, and RMS-amplitude (Pascal) were taken on the first vowels of each word. Spectral tilt was calculated as in Cutler, Wales, Cooper, and Janssen (2007) by comparing the energy in a low frequency band (i.e., up to 400 Hz) relative to that in higher frequency regions. Measurements were based on the vowels rather than on the whole syllables to avoid possible confounds with the number of segments in the syllable for duration measures, and with the presence of unvoiced segments for the other measures.

First, we examined the acoustic dimensions on which the first vowels of the stress pairs differed. Outlier word pairs in which one of the items had values above or below 2.5 standard deviations from the mean on one of the acoustic measures were eliminated. These were: OCtopus - okTOber (pitch), Alibi - aLInea (spectral tilt), Averechts - aveRIJ (spectral tilt), DEcibel - deCIsie (RMS-amplitude). Since three of these four excluded pairs came from the 1-2 stress contrast, the remaining four pairs of this contrast were not analyzed separately but rather pooled with the word pairs from the 1-3 contrast. Paired t-tests showed that words with initial and non-initial stress differed significantly on all four acoustic measures (see Table 4). As expected, vowels of stressed initial syllables were longer, louder, higher in pitch, and had more energy in higher frequency bands. These results were not dependent on the inclusion of the 1-2 word pairs in the analysis. They persisted for analyses of 1-3 word pairs alone. The patterns of the eye-tracking results did not change with these outliers removed.

To examine which of these cues were picked up by the listeners, correlations between the acoustic and behavioral measures were calculated. Correlations were performed on difference measures between initially and non-initially stressed words for the acoustic measures and for the eye-movement data from the critical time window. Measures of non-initially stressed targets were subtracted from measures of initially stressed targets. The results in Table 5 reveal a trend that a greater difference in first vowel duration between items of a pair led to a larger difference in the amount of fixations on the respective targets.

Measurement	stressed vowel (mean)	unstressed vowel (mean)	t (19)	p	Cohen's d
duration (ms)	109	72	5.4	< .001	1.21
mean pitch (Hz)	237	188	17.75	< .001	3.97
spectral tilt	0.8	0.34	2.79	< .01	0.62
RMS (Pascal)	0.08	0.06	2.61	< .01	0.58

Note that spectral tilt is a ratio and therefore does not have a unit.

Table 4. *Mean values of acoustic measures on the first vowel from initially and non-initially stressed words of a pair and significance levels of their difference.*

	target fixations		fixations on competitor - distractors		duration (ms)		mean pitch (HZ)		spectral tilt	
	r(18)	p	r(18)	p	r(18)	p	r(18)	p	r(18)	p
duration (ms)	.43	= .06	.19	= .42						
mean pitch (Hz)	-.15	= .52	-.35	= .13	-.44	= .06				
spectral tilt	-.07	= .77	.34	= .14	.30	= .19	-.33	= .15		
RMS (Pascal)	.27	= .25	-.46	< .05	.22	= .35	.10	= .67	-.36	= .12

Note that spectral tilt is a ratio and therefore does not have a unit.

Table 5. *Correlations among the acoustic measures of the initial/non-initial stress differences and correlations of these acoustic measures with the initial/non-initial difference in target fixations and in fixation differences between competitor and distractors.*

No other correlations with target fixation behavior approached significance. This suggests that vowel duration was the most important cue to a word's stress pattern. The most important cue to influence competition was RMS-amplitude. The difference between initially stressed and unstressed competitors was smaller the clearer the RMS-amplitude stress cues were on the targets.

The acoustic difference measures were entered as predictors in regression analyses. A backward regression model with the fixation difference between initially and non-initially stressed targets as dependent variable showed that the difference in duration remained as the only predictor in the model (R^2 = .182, adjusted R^2 = .136, t(19) = 1.99, p = .06). A backward regression model with the difference of competitor fixations left RMS-amplitude

as the only predictor ($R^2 = .211$, adjusted $R^2 = .167$, $t(19) = -2.19$, $p < .05$). These results support the correlation analyses reported above and confirm that eye-fixation behavior in the critical time window reflects uptake of stress cues.

General Discussion

We investigated the time-course of the use of suprasegmental stress information during spoken-word recognition. We showed that Dutch listeners appear to use all relevant information to recognize words as soon as it comes available. Although other studies have shown that the Dutch can use stress information in making perceptual judgments (Cutler & van Donselaar, 2001; van Donselaar et al., 2005; van Leyden & van Heuven, 1996), none of them demonstrated that listeners efficiently use stress information so early in the recognition process. Eye-tracking allowed us to tap directly into the time-course of processing and show that stress information immediately modulates word recognition.

The time-course of word recognition was as follows. Competition for recognition started as soon as acoustic information about the target came available. 180 ms after target onset, fixations on the segmentally mismatching distractors became less frequent than fixations on target or competitor. After this point in time, there were three primary results. First, information about whether the first syllable had primary stress or not started to modulate the amount of competition. Second, stress contrast (i.e., whether word pairs had stress on the first vs. the second or on the first vs. the third syllable) affected fixations on non-initially stressed targets. Third, before words of a pair became segmentally distinctive, the target was fixated more frequently than its competitor, that is, listeners indeed used stress information to recognize the target.

The first result (i.e., initially stressed words rapidly became stronger competitors than non-initially stressed words) suggests that listeners' preference to hear initially stressed words is at least partially signal driven and not due entirely to the statistical bias in the Dutch lexicon (van Heuven & Hagman, 1988). A related result was found in the correlation analyses between acoustic measures of stress and listeners' eye-movements. A large durational difference on the first vowels of words in a pair facilitated the recognition of initially stressed targets. A small duration difference, however, was more ambiguous in regard to stress in that it did not differentially support initially or non-initially stressed words. Words with ambiguous stress cues might generally be recognized more slowly and suffer from more competition than words with non-ambiguous stress cues. This was also

suggested by Reinisch, Jesse, & McQueen (2008a), who found with similar materials and task that when fewer stress cues were present in the signal, the overall amount of competition was more than when all stress cues were present. In general, therefore, the presence of stress cues appears to be more informative than their absence. Whereas the presence of stress cues tends to enhance the support for initially stressed words, the absence of stress cues tends not be taken as support for the lack of stress. This is probably because in the latter case stress cues could have been reduced (during speech production) or missed (during low-level perceptual processing). The initial stress bias thus appears to emerge at least in part from the continuous uptake of stress cues from the speech signal.

The second result (i.e., the contrast-based asymmetry) also appears to reflect uptake of stress information over time. During the processing of the first syllable, initially stressed targets from both stress contrasts were fixated about equally often. Non-initially stressed targets from the 1-2 contrast, however, received more fixations than non-initially stressed words from the 1-3 contrast. Although primary-stressed initial vowels differed significantly from other initial vowels in all acoustic measures, the secondary stressed initial vowels in the 1-3 stress contrast might be more difficult to distinguish from primary stressed vowels than the unstressed initial vowels in the 1-2 contrast. That is, whereas the first vowels of 'CENtimeter' and 'sentiMENT' both carry some stress, the first vowels of 'OCtopus' and 'okTOber' are, respectively, stressed or unstressed. Likewise, the effect that targets from the 1-2 stress contrast were fixated more frequently than targets from the 1-3 contrast while listeners were processing the second syllable can be attributed to the amount of information conveyed by the second syllable. For 1-2 contrast words, the stress status of the second syllable is informative about the word's identity, because the second syllable of these words is either stressed or unstressed. Both words from a 1-3 stress pair, however, have an unstressed second syllable.

The third result supported our critical hypothesis about the early uptake of phonetic information. Listeners tend to use lexical stress before segmental information could disambiguate the words. The point-by-point analysis showed that looks to the target were more frequent than looks to the competitor before the words could be distinguished by segmental information. In three out of four conditions looks to the target were more frequent than looks to the competitor at or even before the onset of the last matching segment. This demonstrates that listeners use stress information alone to recognize words, rather than segmental cues such as beginning coarticulatory information specifying the segment following the target-competitor divergence point. The exception were words with primary stress on the third syllable. Their stress pattern is the most difficult to recognize

because they have secondary stress on their initial syllables; consistent with our account, these are the items whose recognition should be slowest.

This study has shown that listeners use all relevant phonetic segmental and suprasegmental information to recognize words fast and efficiently. We were able to locate the use of Dutch lexical stress at the earliest moments in the process of word recognition and to attribute performance asymmetries to differences in the information that could be extracted from the speech signal. This investigation of lexical stress provided a good way of asking a more global question about the use of phonetic information. How do listeners cope with the high temporal demands of fast and efficient word recognition? In many cases suprasegmental information may be less informative than segmental information (see, e.g., Cutler & Chen, 1997; Cooper, Cutler, & Wales, 2002), and there listeners may tend to focus on segmental cues. But our results suggest that when suprasegmental information is more useful than segmental information, for example during temporary segmental ambiguities, listeners do use suprasegmental information. Listeners thus use all relevant phonetic information and they do so as soon it comes available.

Appendix

List of stress target pairs grouped by stress contrast and location and their log-transformed CELEX lexical frequency. Stress is indicated in capitals.

	stress location			
	initial		not initial	
1-2 contrast	Alibi (alibi)	2.25	aLInea (paragraph)	2.31
	DEcibel (decibel)	1.15	deCIsie (decision)	0.48
	FYsicus (physicist)	2.03	viSIte (visit)	2.48
	OCtopus (octopus)	1.54	okTOber (October)	3.25
	Opium (opium)	2.37	oPInie (opinion)	2.86
	SYllabus (syllabus)	0.9	syLLAbe (syllable)	0.95
	TErriër (terrier)	1.54	teRRIne (terrine)	1.26
1-3 contrast	Ananas (ananas)	1.99	anaCONda (anaconda)	0.78
	Averechts (contrarily)	2.22	aveRIJ (damage)	1.32
	BArometer (barometer)	1.52	baroNES (baroness)	2.04
	CAvia (guinea-pig)	1.79	kaviAAR (caviar)	2.21
	CENtimeter (centimeter)	3.07	sentiMENT (sentiment)	2.34
	DIameter (diameter)	2.09	diaMANT (diamond)	2.65
	DOminee (pastor)	2.91	domiNANT (dominant)	2.48
	DUbio (doubt)	0.9	dubiEUS (questionable)	2.36
	Ethicus (ethicist)	1.58	etiKET (label)	2.76
	HOSpitaal (hospital)	2.45	hospiTANT (practice teacher)	0
	INdigo (indigo)	2.03	indiGEStie (indigestion)	1.64
	INfanterie (infantry)	1.99	infanTIEL (childish)	2.07
	MEdium (medium)	2.96	mediCIJN (medicine)	3.03
	Opera (opera)	2.66	opeRAtie (operation)	3.23
	RAdius (radius)	0.78	radiAtor (radiator)	1.92
	REquiem (requiem)	1.56	rekwiSIET (stage-property)	1.45
	SPIritus (methylated spirits)	1.95	spiriTIST (spiritualist)	1.49

Lexical stress information modulates the time-course of spoken-word recognition

Chapter 3

Reinisch, E., Jesse, A., & McQueen, J. M. (2008). Lexical stress information modulates the time-course of spoken-word recognition. *Proceedings of Acoustics'08 (CD-ROM,* pp. 3183-3188). Paris: Société Française d'Acoustique.

Abstract

Segmental as well as suprasegmental information is used by Dutch listeners to recognize words. The time-course of the effect of suprasegmental stress information on spoken-word recognition was investigated in a previous study, in which we tracked Dutch listeners' looks to arrays of four printed words as they listened to spoken sentences. Each target was displayed along with a competitor that did not differ segmentally in its first two syllables but differed in stress placement (e.g., 'CENtimeter' and 'sentiMENT'). The listeners' eye-movements showed that stress information is used to recognize the target before distinct segmental information is available. Here, we examine the role of durational information in this effect. Two experiments showed that initial-syllable duration, as a cue to lexical stress, is not interpreted dependent on the speaking rate of the preceding carrier sentence. This still held when other stress cues like pitch and amplitude were removed. Rather, the speaking rate of the preceding carrier affected the speed of word recognition globally, even though the rate of the target itself was not altered. Stress information modulated lexical competition, but did so independently of the rate of the preceding carrier, even if duration was the only stress cue present.

Introduction

When speech unfolds, the incoming speech signal is continuously decoded and evaluated in terms of its similarity to entries in the mental lexicon. The support for a candidate word and hence its role in competing for recognition is dependent on its segmental but also on its suprasegmental similarity to the input. We examine here the role in word recognition of one type of suprasegmental information, namely, lexical stress information. More specifically, we ask whether, duration, as a suprasegmental cue to lexical stress, is evaluated relative to the speaking rate of the preceding context.

Dutch provides a good test case of the use of stress information. Compared to English, Dutch lexical stress is less often marked by vowel quality differences but instead is often only implemented suprasegmentally, that is, through systematic changes in duration, spectral balance, pitch, and amplitude (Sluijter & van Heuven, 1996). Therefore, the Dutch words centimeter and sentiment, for example, share the same segmental beginning /sEnti/, although they differ in stress placement. CENtimeter has primary stress on the first syllable, sentiMENT is stressed on the third syllable, but has secondary stress on the first syllable (syllables with primary lexical stress are marked with capital letters). Corpus studies show that the inclusion of cues to lexical stress information substantially reduces the problem of competition from embedded words in Dutch, but only to a lesser extent in English (Cutler & Pasveer, 2006). Furthermore, Dutch words become lexically distinct earlier if stress is included in the transcription (van Heuven & Hagman, 1988).

Dutch listeners take advantage of stress information to resolve lexical competition during word recognition (van Donselaar, Koster, & Cutler, 2005). However, listeners need to hear the beginning two syllables of a word (e.g., /sEnti/) to inhibit a segmentally mismatching stress competitor (e.g., centimeter for the target sentiment). Stress information on the first syllable (e.g., /sEn/) is not sufficient to suppress the competitor.

Since the suppression of lexical competition appears to depend on the amount of speech material that has been presented, it was of interest to examine the exact time-course of the use of lexical stress in spoken word recognition. The eye-tracking paradigm provides a way of tracking the time-course of word recognition: eye-movement studies have shown that listeners closely follow speech input by looking at related referents presented to them visually (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974). Reinisch, Jesse, and McQueen (2008b) therefore investigated the exact time-course of the use of lexical stress with the eye-tracking paradigm. Listeners' eye-movements to four printed words on a display were tracked as they listened to instructions to click with the computer mouse on

one of them. Critical displays consisted of a stress word pair (e.g., centimeter and sentiment) and a distractor word pair (e.g., alias and alligator). While hearing a stress pair item as a target, listeners looked more often to the target than to the competitor before unique segmental information became available. Some information of the second syllable was, however, needed in order to distinguish the target from its stress competitor. The strength of the competition from the stress mismatching word was modulated by whether or not the target had stress on the first syllable. Target words with primary stress on the first syllable suffered from less competition than words with primary stress on the second or third syllable. In other words, CENtimeter was a stronger competitor for the target sentiMENT than sentiMENT was for the target CENtimeter. One possible explanation is that the presence of stress cues on the first syllable is more salient than the absence of stress cues. In the absence of stress cues on the first syllable, information from the second syllable might be necessary before the competitor can be inhibited. The asymmetry in competition could also be at least partially due to a bias induced by the distribution of stress patterns in Dutch. Most Dutch words have primary stress on their first syllable (van Heuven & Hagman, 1988). Thus, in the absence of sufficient stress information, the stress pair item with word-initial stress is the more likely candidate.

The stress cues that are present on the first syllable of a word, however, are not perceived in an absolute fashion. They are perceived relative to preceding context. The most important cue to lexical stress in Dutch that is independent of sentence accent is duration (Sluijter & van Heuven, 1996). Duration, in turn, is also not perceived in an absolute fashion but rather relative to the rate at which an utterance is spoken. Speaking rate is known to influence the perception of segments with a duration contrast (e.g., Summerfield, 1981). It also modulates the interpretation of duration as a cue to word boundaries (e.g., Repp, Liberman, Eccardt, & Pesetsky, 1978). Duration as a stress cue should therefore also be interpreted in relation to speaking rate.

The purpose of this study was thus to investigate whether speaking rate has an influence on the perception of durational cues to stress and consequently on the resolution of lexical competition during word recognition. The study used eye tracking to investigate the effect of speaking rate in online word recognition. The same materials as in Reinisch, et al. (2008b) were used, but the carrier sentence was either artificially sped up or slowed down. Importantly, the targets themselves were not modified. If changing the speaking rate of the carrier influences the perceived duration then it should in turn modulate the perceived stress pattern of the first syllable of the target word. Syllables will sound longer after fast than after slow contexts. Short syllables (e.g., sen in sentiMENT) will therefore

sound more stressed in fast than in slow contexts. Presented with words with no word-initial stress, participants should look initially more to the word-initial stress competitor (e.g., CENtimeter) in the fast than in the slow context condition. Likewise, long stressed syllables will sound less stressed in slow than in fast contexts. Hearing word-initially stressed targets, listeners should look initially more at the stress competitor with no word-initial stress (e.g., sentiMENT) in slow than in fast context conditions.

Experiment 1

Methods

Participants

24 Dutch native speakers from the participant pool of the MPI for Psycholinguistics were paid for participation.

Stimuli

24 three- and four-syllable Dutch word pairs that overlapped segmentally on their first two syllables but differed in stress position were selected. Seven word pairs had a stress contrast on the first vs. the second syllable (1-2 contrast), for example, OCtopus vs. okTOber. 17 word pairs had a stress contrast on the first vs. the third syllable (1-3 contrast), for example, CENtimeter vs. sentiMENT. Each stress pair was assigned to a distractor word pair that differed segmentally and orthographically from it (e.g., alias and alligator). Distractor word pairs shared the same amount of segmental overlap as the stress pairs but did not necessarily differ in stress location. In addition, eight filler trials were created with two tokens of similar pairs each. Words that appeared together on the screen were semantically unrelated and were as closely equated as possible on their CELEX word frequencies (Baayen, Piepenbrock, & Gulikers, 1995).

A female Dutch native speaker was recorded in a sound attenuated room. Target words were uttered at the end of the carrier sentence "Klik nog een keer op het woord X" ("Click once more on the word X") with sentence accent falling on the target. Multiple recordings of each sentence were made at a neutral speaking rate to facilitate matching the duration of the tokens within each pair.

Speaking rate was estimated based on the duration of the carrier sentence. Note that all carriers had the same content. Speaking rate was changed on the carrier sentence but not on the actual target. The degree to which speaking rate was changed was determined on an item-by-item basis. That is, the speaking rate of the carrier of one item in a stress pair was altered such that the ratio of the duration of the new carrier and the first syllable of the target word was the same as the original ratio for the other stress pair item. For example, the duration of the carrier sentence for CENtimeter was changed such that the ratio of the resulting sentence and the duration of CEN was the same as the ratio of sentiMENT's original carrier sentence and the duration of sen. Similarly, the carrier of sentiMENT was altered so that the carrier to initial syllable ratio matched the ratio of the original CENtimeter sentence to its first syllable.

The sentence durations of seven tokens resulted in extremely fast and slow carrier sentences. Such outliers were defined as tokens for which the new durations were outside the cut-off points of one standard deviation above or below the mean of the (new) sentence durations in the respective speaking rate. The duration of these sentences was set to the values at the cut-off points. For the fast condition, sentences were sped up from 93% to 54% of the original speed (median speed change of 73%). In the slow condition, factors ranged from 100% to 193% (median change of 139%). The distributions in the fast and slow conditions were not overlapping. The durations of the carriers of filler items were assigned randomly but with a similar distribution as those for the target sentences.

All sentences were linearly compressed/expanded using PSOLA algorithm as provided in PRAAT (Boersma, & Weenink, 2007). Although in natural fast or slow speech the duration of segments does not change linearly, this method has the advantage that it does not alter other rhythmic cues in the speech signal (Janse, 2003).

Procedure

Each participant heard half of the words in the fast and half in the slow condition. This assignment was counterbalanced across participants. Trials were blocked by speaking rate. Each display was repeated four times in each speaking rate block. The probability of a target to be the same word as before, its stress competitor, or one of the distractor items was controlled across repetitions. The order of fast and slow blocks as well as the set of words that were targets in the first presentation of a display was balanced across participants. Order of presentation within each block was randomized for each participant. Participants were seated approximately 60 cm in front of a 32.5 by 24 cm screen. The words were

presented in Lucida Sans Typewriter font, font size 20. Eye-movements of the participants were recorded by a head-mounted SMI EyeLinkII eye-tracking system at a rate of 500 Hz. The auditory stimuli were presented over headphones at a comfortable listening level. Participants were instructed to indicate the target word by clicking on the correct item. On each trial, participants saw a fixation cross for 500 ms in the middle of the screen. After 600 ms the words appeared on the screen for 2400 ms. The onset of the presentation of the auditory stimuli was timed such that 1200 ms after the words appeared on the screen the participants heard the onset of the target word. This timing was the same for both speaking rate conditions. The audio signal therefore started before or after the words appeared on the screen, dependent on the duration of the stimulus. This method ensured that participants were given the same amount of time to read the words in all conditions.

Results

Figure 1 shows the mean fixation proportions to target, competitor, and averaged distractors plotted over time from acoustic target onset. Fast speaking rate is indicated by solid lines, slow speaking rate is depicted by dashed lines. The vertical lines show the average syllable offsets per condition shifted by 200 ms, which is an estimate of the time needed to launch an eye-movement related to the acoustic input (Matin, Shao, & Boff, 1993). Figure 1 confirms this assumption: Fixations to the distractors which do not match the segmental acoustic input start to diminish at around 200 ms. The dashed vertical line represents the average segmental Uniqueness Point (UP) of word stress pairs per condition, also shifted by 200 ms.

ANOVAs by participants (F1) and items (F2) were run with speaking rate (fast, slow), stress contrast (first vs. second and first vs. third syllable) and stress location (primary stress on first syllable or not) as factors. Rate was implemented as a within-subject and within-item factor; stress contrast and stress location were within-subject but between-item factors. Repetition, that is whether participants responded to the first or second word of a stress pair, had no effect on responses. This factor was therefore deleted in all subsequent analyses. The ANOVAs were conducted separately for mean proportion of fixations on target and competitor. Mean proportions of fixations on the target were defined as the average ratio of number of fixations on the target compared to all fixations on the four words in the same time window. As a measure of competition, the difference between mean fixation proportions on the competitor and the two distractors was taken. Mean fixation

proportions on competitor and distractors were calculated in the same fashion as for the target, but fixations on the two distractors were averaged.

In a first set of analyses, a time window from 200 to 600 ms after target onset was chosen. This time window encompasses fixations related to the acoustic signal from target onset up to the average segmental UP of target words. Contrary to our predictions, speaking rate did not alter target recognition or the competition process differently for initially stressed and unstressed words (all p's > .05). Speaking rate, however, globally influenced the activation of target words ($F1(1,23) = 4.28$, $p < .05$; $F2(1,44) = 10.02$, $p < .05$): Looks to the target increased earlier when the target followed a carrier sentence presented at a fast than at a slow rate.

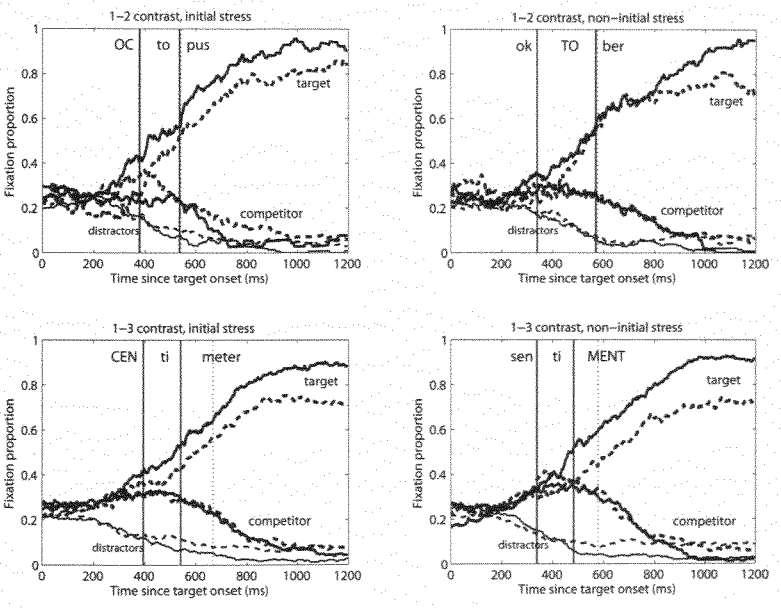

Figure 1. Mean fixation proportions over time to target, competitor, and distractors across conditions. Solid lines represent fast rate, dashed lines slow rate.

Speaking rate of the carrier sentence did not influence the competition process (all p's > .05) but competition was modulated by stress contrast ($F1(1,23) = 6.05$, $p < .05$; $F2(1,44) = 7.40$, $p < .05$): Words from the 1-3 stress contrast competed more strongly for recognition than words from the 1-2 stress contrast. However, stress contrast had no effect on looks to

the target (all p's > .05). Stress location neither influenced fixations on the target nor on the competitor (all p's > .05). None of the interactions was significant (all p's > .05).

To examine the time-course of these effects across syllables, separate analyses on time windows related to the acoustic information of the first and second syllable of each target word were conducted (c.f. the vertical lines in Figure 1). An additional time window encompassed the second syllable plus any further information up to the segmental UPs. Speaking rate did not interact with stress location in any analysis (all p's > .05). An effect of rate on looks to the target was found for all three time windows (first syllable: $F1(1,23) = 5.89$, $p < .05$; $F2(1,44) = 5.66$, $p < .05$; second syllable: $F1(1,23) = 2.96$, $p = .099$; $F2(1,44) = 5.59$, $p < .05$; second syllable up to UP: $F1(1,23) = 3.38$, $p = .08$; $F2(1,44) = 5.94$, $p < .05$). The proportion of looks to the target increased more quickly if the carrier sentence was presented at a fast speaking rate.

At a fast speaking rate participants used stress information to recognize the target. The preference of the target over the competitor was evaluated by taking the ratio of fixations on the target to fixations on target and competitor. This measure indicated a preference for the target while participants were processing the information of the second syllable ($t1(23) = 3.38$, $p < .05$; $t2(47) = 3.05$, $p < .05$). This was not the case for the slow condition (all p's > .05).

Finally, the stress contrast to which a word pair belongs affected how strong the words competed for recognition, especially at the beginning of the competition process. Words from the 1-3 stress contrast competed more strongly than words from the 1-2 stress contrast while the information in the first syllable was processed ($F1(1,23) = 6.22$, $p < .05$; $F2(1,44) = 6.71$, $p < .05$). This effect was not observed in the later time windows.

Discussion

Experiment 1 examined the influence of speaking rate on the perception of duration as a cue to lexical stress during word recognition. Although participants used stress information to distinguish words of a given stress pair, at least when presented after a fast carrier sentence, speaking rate did not modulate stress perception. Rather, speaking rate only had a general influence on target activation in that targets were activated more quickly when preceded by fast than by slow carrier sentences. Note that the duration of the targets themselves had not been altered.

Independently of rate, stress contrast modulated the competition process. Words from the 1-3 stress contrast suffered from stronger competition than words from the 1-2 stress contrast. Words with primary stress on the third syllable have secondary stress on the first syllable. The difference between primary vs. secondary stress is not as salient as between the presence vs. absence of stress. Note that these results are different from Reinisch et al. (2008b), where stress location modulated lexical competition.

One possible explanation for the lack of interaction of speaking rate and stress location in lexical competition could be that duration is not the only cue to lexical stress in Dutch. Other stress cues, that is, spectral balance, pitch, and amplitude, could have cancelled out any effect of speaking rate on the perception of stress location. In the second experiment, pitch and amplitude cues to stress were removed from the first two syllables of the target words.

Experiment 2

Method

24 further participants from the same population used in Experiment 1 were tested. The stimuli from Experiment 1 were modified. Since listeners need more than one syllable for stress perception, pitch and amplitude cues to stress in the first two syllables of each target word were removed. The pitch contours of the first two syllables of each target were measured using PRAAT software (Boersma, & Weenink, 2007) and averaged within a stress pair. Pitch points falling within the first two syllables were subsequently set to the respective average value of the stress pair to generate a flat pitch contour for that part of each target. The pitch manipulation was done in the original context to avoid splicing artifacts. The RMS amplitude of the first two syllables of a word pair was set to the average RMS value of the first two syllables of both words in a pair. The experiment was otherwise identical to Experiment 1.

Results and Discussion

Figure 2 shows mean proportions of looks to target, competitors, and distractors over time. ANOVAs like those in Experiment 1 were run. In the analyses on a time window from 200 to 600 ms after target onset, none of the factors had an effect on mean proportions of

fixations on targets or competitor (all p's > .05). None of the interactions were significant (all p's > .05).

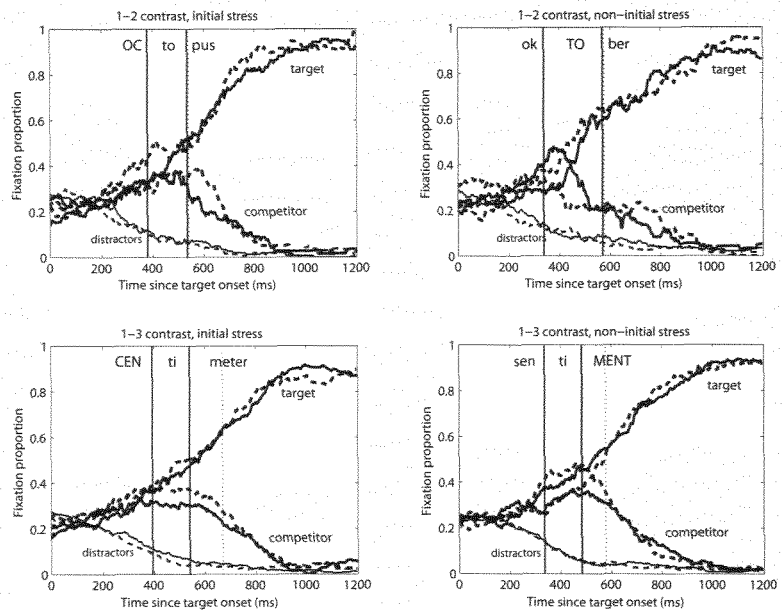

Figure 2. Mean fixation proportions over time to target, competitor, and distractors across conditions. Solid lines represent fast rate, dashed lines slow rate.

Smaller time windows based on the durations of the first and second syllables were then defined. No significant main effects or interactions were found for the time window corresponding to the first syllable duration (all p's > .05). In the time window of the second syllable, speaking rate interacted with stress contrast for looks to the target (second syllable: $F1(1,23) = 6.64$, $p < .05$; $F2(1,44) = 2.60$, $p = .11$; second syllable up to the UP: $F1(1,23) = 7.23$, $p < .05$; $F2(1,44) = 3.56$, $p = .06$). If the carrier sentence was presented at a fast rate, looks to the target were more frequent for words from the 1-3 than from the 1-2 stress contrast. If the carrier sentence was presented at a slow rate, the proportion of looks to the target increased more for words from the 1-2 than from the 1-3 stress contrast.

Participants, however, made use of stress information to recognize the words. They showed a preference for the target before they could have used segmental information to distinguish the words of a stress pair (second syllable: $t1(23) = 2.14$, $p < .05$; $t2(47) = 1.74$, $p = .089$; second syllable to UP: $t1(23) = 2.78$, $p < .05$; $t2(47) = 2.69$, $p < .05$).

To examine whether the removal of pitch and amplitude as stress cues affected the competition process, we combined the data from the two experiments, and added experiment as a between-participant and within-item factor to the analyses. In the time window from 200 to 600 ms after target onset, a main effect of experiment was found ($F1(1,46) = 6.93$, $p < .05$; $F2(1,44) = 9.33$, $p < .05$). In Experiment 2 words competed more strongly for recognition than in Experiment 1. Speaking rate interacted with experiment for target recognition ($F1(1,46) = 5.16$, $p < .05$; $F2(1,44) = 10.31$, $p < .05$). In Experiment 1, looks to the target were more frequent after a fast than a slow carrier. In Experiment 2, this pattern was reversed. In addition, stress contrast affected the strength of lexical competition consistently across experiments ($F1(1,46) = 7.67$, $p < .05$; $F2(1,44) = 9.14$, $p < .05$). Words with a stress contrast on the first vs. the third syllable competed more strongly for recognition than words with a stress contrast on the first vs. the second syllable.

General Discussion

Experiments 1 and 2 investigated whether the speaking rate of a carrier sentence influences the perception of duration in subsequent syllables and hence the perceived stress patterns of these syllables. Word-initially stressed words should compete more for recognition after fast than after slow contexts. Fast carrier sentences should make the first syllable of a word with non-initial stress sound relatively longer and therefore stressed. Slow carrier sentences should change the perceived duration of the first target syllable to sound shorter, thus unstressed. In Experiment 2, pitch and amplitude as a cue to stress were removed from the targets in order to leave duration as the main cue to stress.

Speaking rate, however, did not influence word recognition as a function of whether the first syllable of the target was stressed or unstressed. Speaking rate only had a general effect: In Experiment 1, speaking rate affected the speed at which a target was recognized independently of stress location and stress contrast. In Experiment 2, words from the 1-3 contrast were more quickly recognized after a fast than a slow context. The reversed pattern was found for the 1-2 contrast.

One plausible explanation for the lack of an effect of speaking rate on the perception of stress location could be that listeners need information from more than one syllable to make use of a target word's stress pattern in recognition. If the ratio of the duration of the first and second syllable is considered to determine lexical stress, speaking rate would have less of an effect. In addition, the perceived speaking rate could have been reset, before target word presentation, due to perceptual grouping. The sentences and targets could have been perceived as two perceptual units, since the content of the carrier sentence was the same for all trials and the target word was presented sentence-finally. Consequently, the duration of the target may not have been processed in relation to the speed of the preceding context. Alternatively, since the original durations of stressed and unstressed items were retained, it is possible that the effect of speaking rate was not sufficient to shift the perception of stress categories. If the duration of the syllables had been set to an ambiguous value, then speaking rate might have been able to shift stress perception.

In conclusion, despite the lack of effect of speaking rate on the perception of stress location, participants used stress to resolve lexical competition before words became segmentally unique. Furthermore, duration as a cue to stress was found to be sufficient to induce this effect; although competition was less if other stress cues were also present, stress contrast affected the strength of lexical competition even when only duration was varied. The stress pattern, therefore, influences the time-course of word recognition.

Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue

---

Chapter 4

Reinisch, E., Jesse, A., & McQueen, J. M. (in press). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech*.

Abstract

Three categorization experiments investigated whether the speaking rate of a preceding sentence influences durational cues to the perception of suprasegmental lexical-stress patterns. Dutch two-syllable word fragments had to be judged as coming from one of two longer words that matched the fragment segmentally but differed in lexical stress placement. Word pairs contrasted primary stress on either the first versus the second syllable or the first versus the third syllable. Duration of the initial or the second syllable of the fragments and rate of the preceding context (fast vs. slow) were manipulated. Listeners used speaking rate to decide about the degree of stress on initial syllables whether the syllables' absolute durations were informative about stress (Experiment 1a) or not (Experiment 1b). Rate effects on the second syllable were visible only when the initial syllable was ambiguous in duration with respect to the preceding rate context (Experiment 2). Absolute second syllable durations contributed little to stress perception (Experiment 3). These results suggest that speaking rate is used to disambiguate words and that rate-modulated stress cues are more important on initial than non-initial syllables. Speaking rate affects perception of suprasegmental information.

Introduction

Information about a spoken word is not presented at once but is rather spread over time. Duration is thus an important perceptual cue in speech comprehension, for example to segmental distinctions. When asked to bring a "bin", the person addressed will bring either a bin or a pin depending on how the duration of the voice onset time (VOT) of the initial phoneme is perceived. Durational cues are, however, not perceived in an absolute fashion but rather relative to the rate at which an utterance is spoken. If the sentence "Can you bring me the bin?" is spoken fast, all segment durations will be shorter. Listeners take this into account and interpret duration relative to speaking rate (e.g., Miller, 1981; 1987). If the word "bin" had the same VOT as spoken at a normal rate but followed a faster spoken carrier sentence, it would sound longer relative to the shortened durational cues in the precursor and be more likely to be interpreted as "pin". But duration does not only distinguish words by cueing the identity of their phonemes. It also provides information about a word's suprasegmental structure, for example, its lexical stress pattern. Among other characteristics, stressed syllables tend to be longer than unstressed syllables. Speaking rate should therefore influence the perception of duration as a cue to stress. The present series of experiments investigated whether this is indeed the case and examined the way speaking rate influences the perception of duration as a suprasegmental cue to stress.

Speaking rate has been shown to affect word recognition by influencing the perception of durational cues to phonemic categories such as the perception of VOT in "bin" vs. "pin" (e.g., Allen & Miller, 2001; Miller & Dexter, 1988; Miller & Liberman, 1979; Summerfield, 1981). Likewise, speaking rate can shift the perception of word boundaries. The perception of a word boundary in pairs like 'great ship' vs. 'grey chip' (Repp, Liberman, Eccardt, & Pesetsky, 1978), for example, is cued by the duration of the closure before the fricative as well as the duration of the frication noise. Changes in the perceived relation of these durational cues determine whether the plosive-fricative sequence is interpreted as two distinct phonemes spanning the word boundary or as belonging to a single word-initial affricate (Repp et al., 1978). But speaking rate studies so far have focused on durational cues that concern the perception of one or two segments. Here we test whether speaking rate also affects suprasegmental durational cues. The cues we examined specify a word's lexical stress pattern and are distributed over the syllables of a word.

Dutch provides an ideal test case to investigate these issues. Dutch lexical stress is mostly marked suprasegmentally, that is, by changes in duration, pitch, amplitude, and spectral tilt (Cutler, Wales, Cooper, & Janssen, 2007). Although vowel quality changes,

where unstressed vowels are reduced to schwa, are common in English, they are rare in Dutch. Unlike spectral cues which can change segmental content (i.e., vowel reduction), duration as a cue to stress is expressed on a segment or syllable but does not change the phonemic content. Although lexical stress can be realized by multiple phonetic cues in Dutch, duration is the most important marker of word-level stress. Whereas pitch cues, for example, depend on sentence intonation (Sluijter & van Heuven, 1996), duration cues are partly preserved even if the sentence focus has been shifted to an unstressed syllable (Sluijter & van Heuven, 1995).

The use of suprasegmental stress information in word recognition can be beneficial in Dutch. There are few Dutch word pairs that are distinguished only by lexical stress (e.g., 'VOORnaam', "first name", and 'voorNAAM, "respectable"; capital letters indicate primary stress). Cutler and van Donselaar (2001) found only 13 semantically unrelated pairs. Nevertheless, taking lexical stress information into account substantially reduces the number of embedded words in the lexicon. A lexical entry in Dutch contains on average 1.56 embedded words without stress information considered but only 0.74 embedded words with stress information taken into account (Cutler & Pasveer, 2006). Moreover, lexical stress information shifts segmental uniqueness points considerably closer to the beginnings of words (van Heuven & Hagman, 1988). Without stress information taken into account, Dutch words can be uniquely recognized on average after 80% of their phonemes. With stress information considered, this number reduces to 67% of a word's phonemes.

In line with these computational arguments, Dutch listeners indeed use suprasegmental stress information in spoken word recognition (e.g., Cooper, Cutler, & Wales, 2002; Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005; van Leyden & van Heuven, 1996). Dutch listeners distinguish the first two syllables of word pairs like 'Alibi' and 'aLInea' ("alibi" and "paragraph") by means of their lexical stress patterns (e.g., Cutler & van Donselaar, 2001; van Donselaar et al., 2005). Strikingly, Dutch listeners can do this as soon as the suprasegmental information comes available (Reinisch, Jesse, & McQueen, 2010) and before the words could be distinguished by segmental information. Since duration is a sufficient cue to recognize the stress patterns of words (Reinisch, Jesse, & McQueen, 2008a) speaking rate is expected to influence word recognition by shifting the perception of stress. If the present experiments show that speaking rate influences stress perception, then, given the prior demonstrations of effects of rate on segment identification, the most parsimonious perceptual model would be one with a unitary rate-adjustment mechanism. That is, this mechanism would be concerned not only

with adjusting how segmental information is perceived relative to context, but also with adjusting how suprasegmental information is evaluated.

Unlike local durational cues to segmental categories, the lexical stress pattern of multisyllabic words is distributed over syllables. Furthermore, stress is realized to different degrees on the different syllables of a word. A multisyllabic word has one primary stressed syllable and one or more unstressed or secondary stressed syllables. Note that the degrees of stress on the different syllables in a word are not independent of one another. For example, Dutch words with primary stress on the third syllable (e.g., 'kaviAAR', "caviar") have secondary stress on their initial syllable. Different degrees of stress are realized with different durations (Rietveld, Kerkhoff, & Gussenhoven, 2004; Slootweg, 1988). Primary stressed syllables are longer than secondary stressed syllables which in turn are longer than unstressed syllables. Although listeners can distinguish among these syllables (Mattys, 2000), syllables with secondary stress are more difficult to distinguish from primary stressed syllables than unstressed syllables are to distinguish from primary stressed syllables (Reinisch et al., 2010). The present study therefore investigates whether speaking rate affects the perception of distinct degrees of stress. Since primary and secondary stress are more confusable than primary stress and no stress, speaking rate may be used to a larger extent for the disambiguation of the more confusable contrast. Moreover, given the distribution of stress patterns over the whole word, we ask here whether speaking rate can affect the perception of stress on more than one syllable in a word.

Although stress is distributed over the syllables of a word, previous research on stress perception has shown that in many cases the initial syllable of a word appears to be sufficient for the perception of stress patterns (van Leyden & van Heuven, 1996). Since the initial syllable of a word is directly adjacent to the rate information of the preceding sentence, it was expected that speaking rate would primarily affect the perception of a word's initial syllable as stressed or unstressed. This prediction also follows from the view that spoken-word recognition is incrementally optimal: Information is used as soon as it comes available to inform the word-recognition process (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Norris & McQueen, 2008; Reinisch et al., 2010; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Hence, if speaking rate information and the durational information in the first syllable already strongly determine the word's stress pattern, then we expect that rate will primarily affect the perception of that first syllable.

Speaking rate may nevertheless also affect the perception of the second syllable in a word. First, there may not be enough information in the first syllable to fully specify the

word's stress pattern. Information from the second syllable could therefore still make a contribution to optimal word recognition. Second, speaking rate has previously been shown to affect durational cues that are not immediately adjacent to the rate manipulation. The perception of closure duration as a cue to word medial plosives on a 'rabid' - 'rapid' continuum, for example, depends on the rate of the preceding sentence (Port, 1979; Gordon, 1988). Despite the intervening initial syllable, the plosives were perceived more often as voiceless following a fast than a slow carrier sentence. This suggests that the range of the influence of speaking rate could be large enough to affect the perception of stress on the second syllable of a word.

There is, however, a critical difference between stress perception and the perception of segments. As we have already mentioned, the perception of stress cues on the second syllable is not independent of the perception of stress on the initial syllable. The second syllable could be interpreted in relation to the first syllable and speaking rate may thus affect this relation. If both syllables are perceived as long following a fast context, the perceived durational relation between the first two syllables does not change. Alternatively, however, speaking rate could produce a perceptual chain effect which could cancel or at least diminish the perceived durational difference between the first two syllables, and therefore affect stress perception. That is, following a slow context a stressed initial syllable sounds relatively shorter and, in turn, the second syllable would sound longer relative to the "shortened" initial syllable. This, however, seems unlikely since listeners have no reason to assume a rate change after every syllable. In this study we therefore attempted to control for the dependence of stress on syllables within a word by examining the effect of rate context that is directly adjacent to the second syllable of a word. The initial syllables of the words in the experiment that addresses this issue (Experiment 2) were set to durations that were perceptually ambiguous with regard to stress at a normal speaking rate, and were speeded up or slowed down together with the preceding context. The rate context was thus adjacent to the second syllable while the initial syllable remained ambiguous for stress relative to the preceding context. In this case speaking rate should influence the perception of the second syllable as stressed or unstressed while possible interference from rate effects on the first syllable should be diminished.

The influence of speaking rate on the perception of suprasegmental stress was investigated in a series of two alternative forced choice (2AFC) experiments using the first two syllables of multisyllabic Dutch words. These fragments matched segmentally the beginnings of two longer words which differed in lexical stress placement. 'Alibi' and 'aLInea' have stress on the first vs. the second syllable. There are at least seven other minimal

pairs of this type in Dutch (Reinisch et al., 2010). 'CAvia' ("guinea pig") and 'kaviAAR' ("caviar") have primary stress on the first vs. the third syllable. There are at least 17 other minimal pairs of this type in Dutch (Reinisch et al., 2010). In addition, 'kaviAAR' has secondary stress on the initial syllable. Stress cues other than duration were neutralized. The listeners' task was to decide as quickly and accurately as possible from which of the two words the fragment (i.e., 'ali' or 'kavi') had been taken. The fragments were presented at the end of a fast or a slow version of a carrier sentence. A fast context should make a syllable sound longer and therefore stressed, a slow context should make it sound shorter and unstressed.

These experiments addressed two aspects of spoken-word recognition. First, they sought to specify the nature of the mechanism(s) which adjust for rate variation. As we noted above, the most parsimonious account of an effect of speaking rate context on the perception of stress patterns is that there is a common mechanism for the evaluation of rate information, one that is concerned with the interpretation of both segmental and suprasegmental information. Second, these experiments tested whether spoken-word recognition is incrementally optimal. If so, speaking rate should have stronger effects on how durational information in a word's initial syllable is interpreted than on its second syllable.

Experiments 1a and 1b

Experiments 1a and 1b investigated whether speaking rate affects the interpretation of stress on the initial syllable of the bisyllabic word fragments. Whereas in Experiment 1a the initial syllable was presented at various durations, in Experiment 1b its duration was kept ambiguous for stress. In this way we asked whether the effect of speaking rate would depend on the informativeness of initial syllable duration.

In Experiment 1a the fragments varied along an initial-syllable duration continuum. The second syllable had an ambiguous duration. It was predicted that listeners would perceive the initial syllable durations in relation to the preceding rate context. Moreover, despite the ambiguous second syllables, it was predicted that the perceived initial syllable durations should be sufficient for listeners to recognize the stress patterns of the words. Experiment 1b investigated whether speaking rate affects stress perception when the initial syllable was not informative about stress but durational cues were present on the second syllable. Here, the initial syllable was set to an ambiguous duration while the second syllable

was varied along a duration continuum. Several outcomes were possible. Speaking rate could affect the perception of initial syllable duration and disambiguate it into a relatively long, stressed syllable following a fast context and a relatively short, unstressed or secondary stressed syllable following a slow context. These perceived initial syllable duration differences could be sufficient for listeners to assign the fragments to the words. If, however, listeners also take the duration of the second syllable into account, then the question is whether listeners would do so in relation to the preceding rate context. On this view listeners should be influenced by initial syllable duration relative to the preceding rate context, and by second syllable duration. Hence, they should give more initial stress responses following a fast context than following a slow context and give more initial stress responses the shorter the second syllable.

Experiment 1b addressed a second question about the perception of different stress patterns. For the fragments from 'Alibi' - 'aLInea', where the second syllable is either short when unstressed or long when stressed, the perception of initial stress should decrease the longer the second syllable. It is unclear, however, how listeners would perceive the stress pattern of the fragment from 'CAvia' - 'kaviAAR' if the second syllable was perceived as long. The second syllables of these word pairs are always unstressed and do not naturally differ in duration. Nevertheless, listeners' responses to this second syllable continuum can inform us about the mechanisms of stress perception. If the second syllable becomes longer relative to the first syllable, the shorter first syllable might be perceived as not having primary stress. Alternatively, listeners might show a tendency to judge a long syllable on the fragment as having initial stress irrespective of which syllable was long. That is, they could use overall fragment duration rather than initial syllable duration to assign the fragment to the respective word. A fragment with only two short syllables would then be assigned to the word with primary stress outside that fragment, that is, to the word with primary stress on the third syllable.

In summary, Experiments 1a and 1b investigate the effect of speaking rate on the perceived degree of stress on the initial syllable whether initial syllable duration is informative about stress or not. Further, they address the use of second syllable duration in the perception of stress patterns.

Method

Participants

Fourteen Dutch native speakers from the Max Planck Institute participant pool took part for a small payment. They all reported no hearing disorders. A week after completing Experiment 1a they returned for Experiment 1b.

Materials and Design

Two word pairs of at least three syllables length were selected such that both words of a pair overlapped segmentally on their first two syllables but differed in lexical stress placement. 'Alibi' - 'aLInea' ("alibi" - "paragraph") had primary stress on the first vs. the second syllable (1-2 stress contrast). 'CAvia' - 'kaviAAR' ("guinea pig" - "caviar") had primary stress on the first vs. the third syllable (1-3 stress contrast). The word with primary stress on the third syllable ('kaviaar') had secondary stress on the initial syllable. Note that, despite the orthographic difference, the first two syllables of 'cavia' and 'kaviaar' share the same phonemes. The first two syllables of the words served as fragments that were subjected to the experimental manipulations. For better comparison of durational effects across stress contrasts, word fragments with the same vowels were chosen. Word fragments from multisyllabic words rather than bisyllablic segmental homophones (i.e., words that differ only in stress placement) were used to avoid confounding the effects of duration as a cue to stress with the effects of word-final lengthening (Nooteboom & Doodeman, 1980), and to be able to test different stress contrasts. The words had comparable CELEX lexical frequencies (Baayen, Piepenbrock, & Gulikers, 1995; log-transformed frequencies: Alibi 2.25, aLInea 2.31, CAvia 1.79, kaviAAR 2.21).

The stimuli in all experiments were resynthesized with MBROLA using the diphone inventory of a Dutch female voice. The resynthesis was based on recordings of a Dutch female native speaker who uttered the four words at the end of the carrier sentence 'Klik nog een keer op het woord' ("Click once more on the word"). Word fragments were modeled on the segments of the first two syllables of the words, 'ali' and 'kavi', respectively. They were given a flat pitch contour at the value of the last pitch point in the carrier sentence (190 Hz). The pitch contour in the word fragments was therefore not informative about stress location, neither due to a pitch jump relative to the preceding sentence, nor in its movement within the two syllables. Two versions of the carrier sentence were created to provide a fast and a

slow rate context. Segment durations and pitch contours of the sentences were modeled on one token of the recorded sentences. Expansion and compression of speaking rate was implemented linearly by multiplying each of the segment durations in the natural sentence with .66 for a fast version and with 1.33 for a slow version. The same precursor sentences (i.e., fast and slow) were used for both stress contrasts.

In Experiment 1a the duration of the first vowels of the fragments were varied on 13-step duration continua. Since stress mainly affects vowels (Fry, 1955) only durations of vowels were manipulated. Consonants were set to the average of their original durations in the two words of a pair ('ali': [l] 50 ms; 'kavi': [k] 89 ms, [v] 76 ms). Duration values of the vowels were based on a larger data set of similar word pairs described in Reinisch et al. (2010; see Table 1 for a summary of the duration distributions on initial and second vowels in this data set). The lowest values of the continua were set to the values of the appropriate shortest unstressed first vowel in this larger data set. The longest values of the continua approximately matched the appropriate longest stressed first vowel. Only words from the same respective stress contrast as the manipulated fragment were taken into account. The continua spanned durations from 22 ms to 162 ms for 'ali' (1-2 contrast) and from 34 ms to 174 ms for 'kavi' (1-3 contrast). Steps were separated by 11 ms, except for the endpoints which were 15 ms longer or shorter than their adjacent steps. The durations of the second vowels were set to the average duration of second vowels in the same data set (93 ms for 'ali' and 68 ms for 'kavi').

In Experiment 1b, 13-step duration continua were created based on the second vowels of the fragments. Durations were increased in steps of 11 ms from 38 ms to 170 ms for the [i] in 'ali' (1-2 contrast) and from 24 ms to 156 ms for the [i] in 'kavi' (1-3 contrast). Again vowel durations were modeled on the longest and shortest stressed and unstressed vowels in the larger data set. For both fragments the short endpoint of the continuum was set to a duration close to the shortest duration of an unstressed second vowel within word pairs of the same stress contrast. The long endpoints of the continua were established by adding 11 ms 12 times to create the 13-step continua. Note that the long endpoint for the 1-2 contrast was shorter than that of the longest stressed second vowel in the data set (170 ms vs. 217 ms). This endpoint was chosen because 217 ms was an outlier. For the 1-3 contrast the duration of the short endpoint of the continuum (24 ms) was shorter than the shortest second vowel in the larger data set (33 ms). The long endpoint (156 ms) of the continuum was longer than the longest vowel duration of a second syllable in the data set (112 ms). These values were chosen because words from the 1-3 contrast always have an unstressed, thus short, second syllable. The continuum of the 1-3 contrast should, however, span the

same duration range as the continuum for the 1-2 contrast. The vowel durations of the initial syllables were based on the perceptual data from Experiment 1a. They corresponded to the step of each continuum that was perceived as most ambiguous with respect to stress (i.e., the step at which initial and non-initial responses were given about equally often). For 'ali' this corresponded to an [a] of 92 ms and for 'kavi' to an [a] of 104 ms. These values were close to the average durations of the stressed and unstressed first vowels in the larger data set (see Table 1).

Each of the generated 52 fragments (2 Experiments x 2 Stress contrasts x 13 Steps) was combined with both the fast and the slow version of the carrier sentence. Listeners were presented with all stimuli 10 times. Presentation order was randomized separately for each participant within each block such that all combinations of step and rate would be presented once before a repetition occurred. Fast and slow sentences were mixed at random. Presentation was blocked by stress contrast ('ali' and 'kavi' trials) with a switch of contrast after five repetitions. This allowed listeners to answer to one contrast at a time so that they would not respond contrastively. The order of blocks was counterbalanced across participants. In total each participant responded to 520 stimuli (2 Stress contrasts x 2 Rates x 13 Steps x 10 Repetitions) in each of the experiments.

		Mean duration		Standard deviation		Minimum		Maximum	
stress contrast	stress location	1st vowel	2nd vowel	1st vowel	2nd vowel	1st vowel	2nd vowel	1st vowel	2nd vowel
1-2	1st syllable	107.1	64.4	47	22	53	40	164	99
	2nd syllable	62.3	122.3	28	49	22	86	101	212
1-3	1st syllable	114.6	69.5	38	22	55	34	177	108
	3rd syllable	73.8	65.8	21	22	34	33	132	112

Table 1. *Distributions of initial and second syllable vowel durations (ms) for the different stress contrasts and stress locations in the data set described in Reinisch, Jesse, & McQueen (2010). These values were used as reference points for creating duration continua in the present study. Values for the 1-2 contrast are based on seven word pairs; values for the 1-3 contrast are based on 17 pairs. Durations are given in milliseconds.*

Procedure

Listeners were tested individually in a sound-attenuated booth. On every trial two response alternatives ('alibi' - 'alinea' or 'cavia' - 'kaviaar' respectively) were presented on a screen with the letters that corresponded to the phonemes of the fragment marked in red. The initially stressed word was always presented on the left side of the screen and corresponded to the left button. After 200 ms listeners heard the carrier sentence immediately followed by the word fragment. Stimuli were presented over headphones at a comfortable listening level. The response alternatives stayed on the screen throughout the trial. The listeners' task was to indicate by button press as quickly and accurately as possible from which of the two words the fragment had been taken. Listeners were not informed that the critical question was about the perception of stress placement. The next trial started 3000 ms after target onset or 500 ms after the participant's response. No feedback was given. Every 65 trials participants were given a short break. The experiment was controlled by NESU experimental software (http://www.mpi.nl/world/tg/experiments/nesu.html).

Analysis

Analyses were run separately for each experiment on responses to the fragments from the 1-2 ('ali') and the 1-3 contrast ('kavi'). Data were analyzed using linear mixed-effect models (Baayen, Davidson, & Bates, 2008) as provided in the lme4 package (Bates & Sarkar, 2007) in R (version 2.8.0; 20-10-2008) with response (i.e., initial stress or not) as dichotomous dependent variable and rate (fast vs. slow), step (numerical values 0 to 12), and block (first vs. second; recoded to the numerical values -1 and 1) as fixed factors. Block was defined for each stress contrast as the first five repetitions of all sentence-fragment combinations vs. the second five repetitions of all sentence-fragment combinations. The variable participant was entered as random effect in the models. This allowed the intercept to vary by participant with the restriction that the mean of this random variation was zero. A logistic linking function was used to deal with the categorical nature of the dependent variable. The model maps one condition on the intercept (e.g., fast rate at step 0 for the hypothetical block 0) and assigns regression weights for the adjustments required to map from this condition to every other level of each factor. For the categorical factor rate (i.e., fast or slow) the estimated weight gives the mean adjustment of the intercept from fast to slow rate. Step and block were defined as numerical factors. Here, the regression weight has to be multiplied by the numerical value of the factor to adjust the value that has been mapped on the intercept (value 0) to the respective other levels of the factor. A significant effect can be inferred if a

regression weight is statistically different from zero. The sign of the weight indicates the direction of the change. In our analyses a positive weight means more initial stress responses, a negative weight means fewer initial stress responses. The analysis started out with a full model that included all three factors and their interactions. Non-significant interactions were eliminated from the model if the simpler model fitted the data better than the model including the interaction. After removing the interactions, non-significant factors were eliminated by the same procedure. During this model-fitting procedure, models were tested against each other by means of log-likelihood ratio tests. Only significant results will be reported.

Results

Experiment 1a

Figure 1 shows categorization data in response to the fragments from the 1-2 and 1-3 contrast in Experiment 1a. For the fragment from the 1-2 stress contrast ('ali') main effects of rate, step, and block were found (b_{rate}=-.43; p<.001; b_{step}=.3, p<.001; b_{block}=-.83, p<.001) as well as an interaction between step and block ($b_{step*block}$=.13, p<.001). Listeners gave more initial stress responses at a fast than at a slow rate context. More initial stress responses were given the longer the first syllable duration and there were more initial stress responses in the first than in the second block. Step had a larger effect in the second block than in the first block, that is, listeners' responses became more categorical in block 2.

For the fragments from the 1-3 stress contrast ('kavi'), main effects of step and block (b_{step}=.55, p<.001; b_{block}=-.76, p<.001) and an interaction between these two factors ($b_{step*block}$=.11, p<.001) were found. As for 'ali', listeners gave more initial stress responses the longer the first syllable duration and gave more initial stress responses in the first than in the second block. Listeners' responses again became more categorical in the second block. Although there was no main effect of rate, it interacted with step ($b_{step*rate}$=-.08, p<.01). A post hoc test showed that rate affected the longer initial syllable durations (short durations, steps 0-5: b_{step}=.60, p<.001, b_{rate}=-.13, p=.32; long durations, steps 6-12: b_{step}=.35, p<.001, b_{rate}=-.58, p<.001).

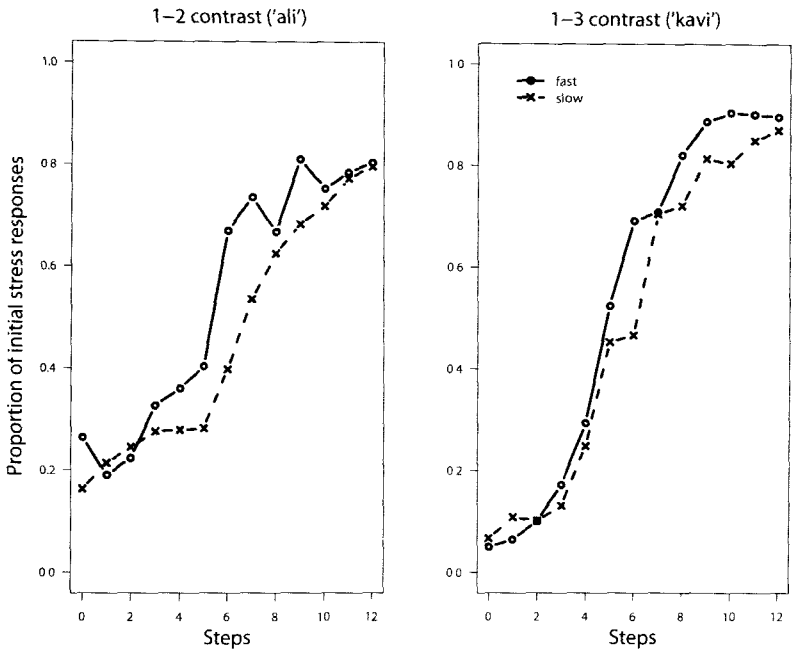

Figure 1. *Proportion of initial stress responses in Experiment 1a for the fragments from the 1-2 ('ali') and 1-3 ('kavi') stress contrast along the duration continuum on the vowel of the respective first syllable. Second syllables were set to the average duration of stressed and unstressed second syllables of tokens from the respective stress contrast.*

Experiment 1b

Figure 2 shows the categorization data for the fragments from the 1-2 and 1-3 stress contrast in Experiment 1b. For the fragment from the 1-2 contrast ('ali') only a main effect of rate was found (b_{rate}= -.82; p<.001). Listeners gave more initial stress responses at the fast than at the slow rate. For the fragment from the 1-3 contrast ('kavi'), rate, step, and block had a significant effect (b_{rate}=-.40; p<.001; b_{step}=.05; p<.001; b_{block}=-.18, p<.01). More initial stress responses were given at a fast than at a slow rate. Step and block also showed a significant interaction ($b_{step*block}$=.03, p<.001). More initial stress responses were given the longer second syllable duration and more initial stress responses were given in the first block. The effect of step was larger in the second block than in the first block, that is, listeners' responses once again became more categorical later in the experiment.

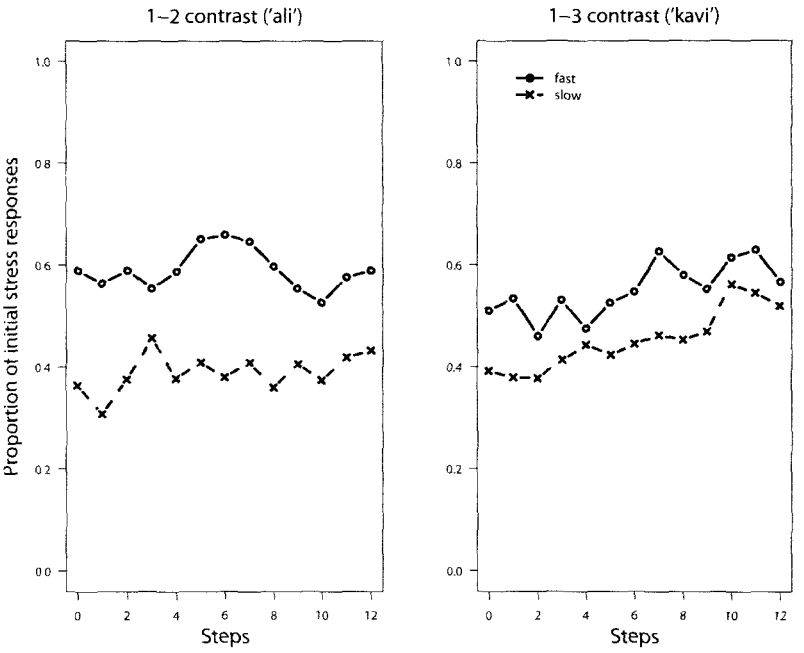

Figure 2. *Proportion of initial stress responses in Experiment 1b for the fragments from the 1-2 ('ali') and 1-3 ('kavi') stress contrast along the duration continuum on the vowel of the respective second syllable. Initial syllables were set to durations that were perceived as ambiguous for stress in Experiment 1a.*

Cross-experiment comparison

Additional analyses were carried out in order to compare the effect of first and second syllable durations on the perception of stress. Mean proportions of initial stress responses were calculated for each participant at the endpoints of each duration continuum pooled over rate. Based on the overall proportions at the endpoints, difference scores were calculated for each participant such that the endpoint with the overall lower mean proportion of initial stress responses was subtracted from the endpoint with the overall higher mean proportion of initial stress responses. The difference scores hence indicate the extent of use of syllable duration. Paired *t*-tests showed that for both stress contrasts the difference in initial stress responses between the endpoints of the duration continua was significantly larger in Experiment 1a than in Experiment 1b (1-2 contrast: $t(13)=4.11$, $p<.001$;

1-3 contrast: $t(13)=8.5$, $p<.001$). Initial syllable duration was used more in the perception of stress location than second syllable duration.

Discussion

Experiments 1a and 1b showed that the speaking rate of the preceding context affects the perception of duration as a suprasegmental cue to lexical stress. The perception of stress location was influenced by the rate whether initial syllable duration was informative about stress (Experiment 1a) or not (Experiment 1b). Following a fast context the initial syllable of the fragment should sound relatively longer, and therefore more stressed than following a slow rate context. Indeed, listeners were more likely to report initial stress when the context was fast than when it was slow. In Experiment 1a listeners considered the steps of the duration continuum in relation to the speaking rate of the preceding context rather than basing their responses on the absolute duration of the syllables alone For the 'ali'-fragments rate affected mostly the middle of the duration continuum (steps 3-10; see Figure 1), while for the 'kavi'-fragments the rate effect appears to be stronger for the longer initial vowel durations (steps 5-12). In Experiment 1b the vowel of the initial syllable of each fragment was set to a duration that appeared to be ambiguous for stress in Experiment 1a (i.e., listeners gave approximately 50% initial stress responses at the respective steps of the continua). Listeners in Experiment 1b, however, did not perceive these first syllable durations as completely ambiguous. Rather they interpreted them as long or short in relation to the preceding rate context. Following a fast rate context the initial syllable was perceived as long and therefore stressed; following a slow rate context it was perceived as short and unstressed. The duration of the second syllable (i.e., the different durations of the continuum) contributed little to listeners' responses. Duration variation on the second syllable affected perception only of the 1-3 contrast fragments and this effect was small in comparison to the effect of initial syllable duration in Experiment 1a. The endpoints of the continuum in Experiment 1b received only 45% and 54% initial stress responses as compared to 6% vs. 88% in Experiment 1a. It thus appears that the duration of the initial syllable together with the information from the rate context was largely sufficient to decide about the stress pattern of the word.

The small effect of second syllable duration for the 1-3 contrast, however, provides additional insight in the perception of stress placement. Listeners gave more initial stress responses the longer the second syllable. This suggests that for fragments from the 1-3 contrast a longer duration of either the first or second syllable was perceived as coming from

a word with initial syllable stress. That is, listeners were more likely to perceive initial stress if the fragment contained at least one long syllable. It is only when listeners heard two short syllables that they judged the stress to fall outside the fragment, that is, on the third syllable of the word. Listeners seem to have considered the overall duration of the fragment to decide which word the fragment was taken from. A long fragment was assigned to the initially stressed word whereas a short fragment was assigned to the word with third syllable stress.

Experiments 1a and 1b showed that speaking rate affects the perceived duration of the initial syllable and thereby influences the perception of suprasegmental lexical stress patterns. The second syllable durations of the continuum contributed little to the perception of stress location even if the initial syllable did not carry cues to stress. Nevertheless, speaking rate could in principle affect the perception and use of second syllable duration. Since rate affected the perception of the presumably ambiguous initial syllables in Experiment 1b the use of the duration continuum on the second syllable may not be necessary for rate to affect its perceived duration. Experiment 2 tested whether this was the case. To assess the effect of speaking rate on the second syllable independently from the effect on the initial syllable, the ambiguous initial syllables of Experiment 1b were included in the rate manipulations. That is, the initial syllables were presented at the same rate as the preceding context and therefore were ambiguous with regard to stress within each speaking rate context. Moreover, we asked whether the absence of stress information on the initial syllable would enhance the use of absolute second syllable duration in stress perception. Since the initial syllable duration was not informative about stress and the second syllable was immediately adjacent to the rate context, an effect of second syllable duration as well as an influence of speaking rate on the perception of this duration were expected. Experiment 2 hence investigated whether speaking rate can affect the perception of suprasegmental lexical stress on the second syllable of a word.

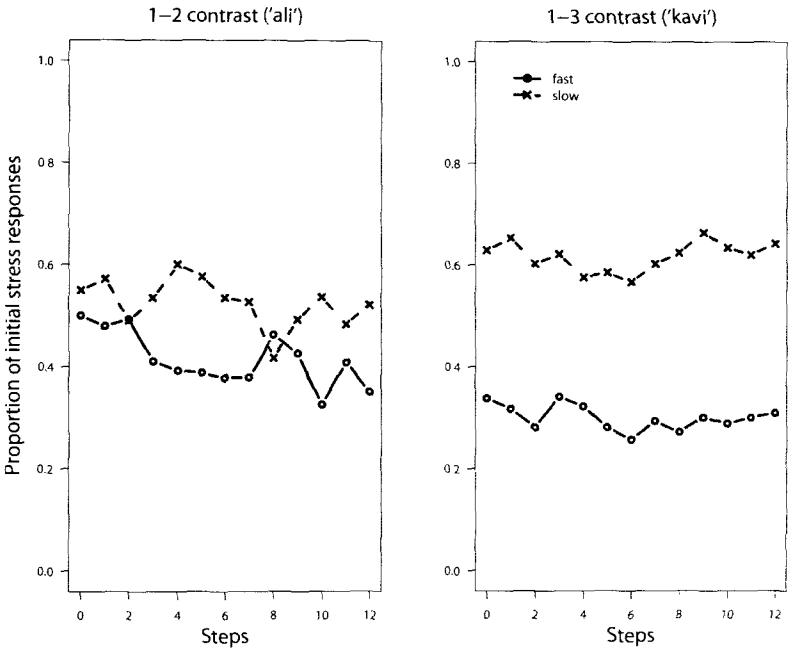

Figure 3. *Proportion of initial stress responses in Experiment 2 for the fragments from the 1-2 ('ali') and 1-3 ('kavi') stress contrast along the duration continuum on the vowel of the respective second syllable. Initial syllables were set to durations that were perceived as ambiguous for stress in Experiment 1a and included in the rate manipulation.*

Experiment 2

Method

Participants

Fourteen participants from the same population as in Experiments 1a and 1b were paid for taking part. None of them had participated in the previous experiments.

Materials, Design, and Procedure

The stimuli from Experiment 1b were manipulated such that the first vowel of each word fragment was speeded up or slowed down at the same rate as the precursor sentence. That is, for the fast context the duration of the first vowel was multiplied by .66; for the slow context it was multiplied by 1.33. The resulting first vowel durations were 61 ms and 122 ms for the 1-2 contrast ('ali') and 69 ms and 138 ms for the 1-3 contrast ('kavi'), respectively. Note that these durations would approximately correspond to steps 3 and 9 of the initial-syllable duration continua (steps from 0 to 12) in Experiment 1a. They were located symmetrically around the middle of the continuum. The duration continua on the second syllable were the same as in Experiment 1b.

The design, experimental setup, task, and analyses were the same as in Experiments 1a and 1b.

Results

Figure 3 shows the proportion of initial stress responses along the continuum for fast and slow rate contexts. For the fragment 'ali' from the 1-2 contrast, main effects of rate and step were found (b_{rate}=.47; $p<.001$; b_{step}=-.03, $p<.01$). Initial stress responses were more frequent after a slow than after a fast context and at shorter second syllable durations. For the fragment from the 1-3 contrast ('kavi') only the factor rate (b_{rate}= 1.35, $p<.001$) was significant. More initial stress responses were given following a slow than a fast context.

Discussion

Listeners again used rate information to perceive the suprasegmental stress pattern of words from both stress contrasts. More initial stress responses were given when the second syllable followed a slow than a fast context. A slow context made the second syllable sound shorter and therefore unstressed. Speaking rate thus affects a syllable that is adjacent to the rate manipulation independent of whether stress cues on that syllable are used or not. The duration of the continuum on the second syllable had little effect on the perception of stress. Unlike in Experiment 1b, however, where a small effect of second syllable duration was found for the 1-3 contrast, here an effect was found for the 1-2 contrast. This suggests that the duration of the second syllable contributes to the perception of primary stress location at least to some extent.

Experiments 1b and 2 support previous findings that the initial syllable is the most important syllable in the perception of Dutch stress patterns (e.g., van Heuven & Hagman, 1988). Note that the absence of any strong effects of the second syllable continuum was not specific to one stress contrast. It is therefore worth asking why listeners did not use much information about the second syllable. Given that in this task listeners gave their response after the offset of the stimuli, it is unlikely that listeners made their decisions about stress location before information from the second syllable came available. But it is possible that listeners might not have perceived the variability in the duration of the fragment-final second vowel. The end of continuous sounds like vowels may be less easily detected when followed by silence than by further speech context. Experiment 3 tested whether this could have been the case by replicating Experiment 1b with an additional [t] that clearly marked the end of the fragments with its burst.

Experiment 3

Method

Participants

Fourteen participants who had not taken part in any of the previous experiments received a small payment for their services. None of them reported any hearing problems.

Materials, Design, and Procedure

The sound [t] was added to the stimuli used in Experiment 1b since its burst clearly marks the end of the fragment. The [t] was synthesized with a duration of 100 ms and without intervening silence, so that it was coarticulated with the word-fragments. The sound [t] is not a possible continuation of any of the fragments used here that forms an existing Dutch word. Listeners were informed that they would hear the word fragments 'ali' and 'kavi' to which a [t] had been added (i.e., 'alit' and 'kavit') and that they should decide which word on the screen the fragment was "taken from". The experimenter suggested that the [t] should be ignored. The task was thus similar to gating tasks where fragments are frequently followed by noise (e.g., Smits, Warner, McQueen & Cutler, 2003). Procedure and design were otherwise the same as in the previous experiments.

Results

Figure 4 shows categorization data for the fragments 'ali' and 'kavi' in Experiment 3. For 'ali' (1-2 contrast) a significant main effect of step (b_{step}=-.09, p<.001) was found. More initial stress responses were given the shorter the second syllable durations were. Furthermore, the three-way interaction between block, step, and rate ($b_{rate*step*block}$=-.04, p<.05) was significant. Subsequent analyses of this interaction for each block revealed a main effect of step in the first block (b_{step}=-.08, p<.001; more initial stress responses the shorter the second syllable) and an effect of rate and step in the second block (b_{rate}=-.11, p<.001; b_{step}=-.33, p<.001; more initial stress responses the shorter the second syllable and following a fast context). For the fragment from the 1-3 contrast ('kavi') significant main effects of rate, step, and block were found (b_{rate}=-.58, p<.001; b_{step}=.1, p<.001; b_{block}=-.19, p<.01). Listeners gave more initial stress responses at a fast than at a slow rate. Step and block also showed a significant

interaction ($b_{step*block}$=.03, p<.001). More initial stress responses were given the longer the second syllable durations were, and more initial stress responses were given in the first block. The effect of step was larger in the second block, that is, responses became more categorical in the course of the experiment.

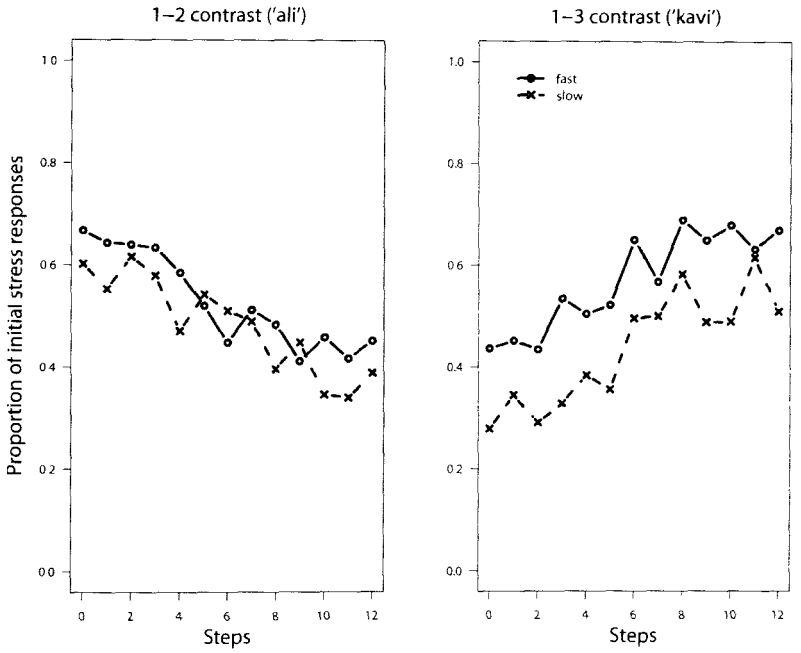

Figure 4. *Proportion of initial stress responses in Experiment 3 for the fragments from the 1-2 ('ali') and 1-3 ('kavi') stress contrast along the duration continuum on the vowel of the respective second syllable followed by a sound [t]. Initial syllables were set to durations that were perceived as ambiguous for stress in Experiment 1a.*

Cross-experiment comparison

Experiment 3 was compared to Experiment 1b to test whether the addition of the [t] as a marker of the end of the fragment increased the use of second syllable duration. As for the comparison of Experiments 1a and 1b, difference scores between the proportions of initial stress responses at the endpoints of the duration continua were calculated for each participant pooled over rate. The endpoint with the lower overall mean proportion of initial stress responses was subtracted from the endpoint with the overall higher mean proportion of initial stress responses. Independent sample t-tests showed that in both stress contrasts, second syllable duration was not used more often in the decision about lexical stress location in Experiment 3 (i.e., after the addition of the [t]) than in Experiment 1b (1-2 contrast: $t(26)=1.22$, $p=.23$; 1-3 contrast: $t(26)=.92$, $p=.37$).

Discussion

Experiment 3 showed that the weak effects of second syllable duration found in Experiment 1b and 2 were not due to the fact that listeners did not perceive the duration differences in fragment-final position. Although listeners used durational information of the second syllable to decide about the stress pattern of the fragments at least to some extent, they did not do so more when the end of the fragments was marked by the burst of a [t] than when it was followed by silence.

The directions of the effect of second syllable duration for the two stress contrasts were consistent with the effects found in the previous experiments. Whereas for the 1-2 contrast listeners gave more initial stress responses the shorter the second syllable (as in Experiment 2), for the 1-3 contrast more initial stress responses were given at longer second syllable durations (as in Experiment 1b). Although the response pattern for the 1-3 contrast might have been unexpected in Experiment 1b, its replication in Experiment 3 suggests that it did not occur by chance. If the second syllable duration is expected to be uninformative about the stress pattern (i.e., words with primary stress on the first or the third syllable both have a short second syllable) listeners appear to rely on the whole fragment duration. They assigned fragments with a long syllable to the initially stressed word irrespective of whether it was the first or second syllable. Only two short syllables led to the perception of primary stress outside the fragment, that is, stress on the third syllable.

Speaking rate affected the perception of stress such that more initial stress responses were given following a fast than a slow context. For the 1-2 contrast, however, this effect was

restricted to the second block. This could have occurred due to task difficulties. Participants reported after the experiments that they found it difficult to ignore the following [t]. With their attention focused on the [t] earlier rate information could have been missed. Note that the [t] could not have provided rate information that attenuated the effect of the preceding context since the duration of the [t] was kept constant across the experiment.

Experiment 3 indicates that listeners can use second syllable duration to decide about the stress patterns of the word fragments. The relatively weaker use of second syllable duration in the perception of stress location as compared to the use of initial syllable duration in Experiments 1a and 1b, however, can not be explained by listeners' failure to perceive the end of the second syllable in fragment final position.

General Discussion

In a series of categorization experiments we examined whether and how the speaking rate of a preceding sentence context influences the perception of durational cues to the lexical stress patterns of Dutch multisyllabic words. Previous research on speaking rate has focused mainly on rate effects on the perception of segments. Dutch lexical stress, however, is almost exclusively marked suprasegmentally. So Dutch provided a good test of whether speaking rate influences the perception of a purely suprasegmental distinction. Moreover, unlike local durational cues to phonemes or word boundaries, lexical stress patterns of multisyllabic words are distributed over more than one syllable. We therefore asked whether speaking rate can affect stress perception on initial and non-initial syllables and how speaking rate affects the perception of different degrees of stress in different stress patterns (i.e., 1-2 and 1-3 contrasts).

Results showed that speaking rate affected the perception of suprasegmental stress on initial syllables of Dutch word fragments from both tested stress contrasts (the 1-2 contrast and the 1-3 contrast). Since all phonetic information is used as soon as it comes available (Reinisch et al., 2010), speaking rate was expected to immediately exhibit its effect on the perception of initial syllable stress. Following a fast context sentence the initial syllable should be perceived as longer and therefore more stressed than following a slow context sentence. Experiments 1a and 1b showed that speaking rate is not only used to disambiguate syllables that are ambiguous for stress. Speaking rate also contributes to the perception of syllables that could have been assigned to stress patterns on the basis of their absolute durations. Furthermore, similar effects for both stress contrasts showed that

speaking rate was used to distinguish different degrees of stress. Primary stressed syllables could be distinguished from unstressed syllables (1-2 contrast) as well as from secondary stressed syllables (1-3 contrast).

Speaking rate also affected the perception of stress on the second syllable. This is remarkable since listeners hardly used the duration continuum on the second syllable to decide about the fragments' stress patterns. As pointed out in the introduction, the stress status of a syllable in a word is not independent of the presence of stress on other syllables. We therefore restricted our investigation to immediately adjacent rate context on the second syllable. That is, the initial syllables of the fragments were made ambiguous for stress and manipulated along with the rate contexts. Experiment 2 showed that despite listeners' lack of attention to the second syllable duration continuum, speaking rate influenced and thus disambiguated second syllable durations. This effect was similar to what was observed with ambiguous initial syllable durations in Experiment 1b. There is, however, an alternative explanation for this rate effect on the second syllable. Rather than using speaking rate to decide whether the second syllable was stressed or unstressed, listeners could have relied on the duration of the initial syllable to decide about the fragments' stress patterns. Note that if the initial syllable was included in a fast context it was short whereas in a slow context it was long. Listeners might not have perceived the initial syllable as included in the rate context and thus not as ambiguous for stress. This use of initial syllable duration might have led listeners to neglect stress information on the second syllable.

In addition to the main question about speaking rate, the investigation of the 1-3 stress contrast also addressed how listeners perceive stress cues on different syllables. The second syllable of both words from the 1-3 stress contrast is always unstressed, hence expected to be always short. If fragments from the 1-3 contrast were presented with long second syllables, listeners surprisingly but consistently (in Experiments 1b and 3) reported word initial stress. They appeared to use overall fragment durations rather than second syllable durations to interpret the stress patterns. This finding is novel and could not have been found in previous studies on lexical stress perception since these used mostly bisyllabic words (i.e., only stressed vs. unstressed syllables) and, in addition, varied initial and second syllable duration at the same time (e.g., Fry, 1955; van Heuven & Menert, 1996).

Even though the second syllable duration contributed to stress perception to some extent in both stress contrasts, the results emphasized the importance of the initial syllable in stress perception. Experiment 3 thereby replicated Experiment 1b by demonstrating that it was not the failure to perceive the end of the fragment that led to the rare use of second

syllable duration. Whenever the duration of the initial syllable of the fragment could be interpreted in its degree of stress, the second syllable duration continuum was given little attention. This is in line with previous findings which also showed that the initial syllable of a word is largely sufficient for listeners to use lexical stress in word recognition (e.g., van Leyden & van Heuven, 1996). This suggests that lexical stress patterns are not perceived solely as duration ratios between initial and second syllables. Rather, stress patterns are perceived incrementally as the signal unfolds (see also Reinisch et al., 2010). As with segmental information (e.g., Dahan et al, 2001), therefore, suprasegmental information is taken into account as soon as it comes available. The current results thus support the view that spoken-word recognition is incrementally optimal (Norris & McQueen, 2008; Tanenhaus et al., 1995), where all sources of information are used as fully as they can be and as early as possible.

Listeners interpret suprasegmental stress patterns relative to speaking rate information in the preceding context in a similar fashion to how they deal with rate information in making segmental distinctions. Given this similarity between segmental and suprasegmental processing, it is reasonable to assume that there is a common mechanism for the evaluation of both kinds of cues, both locally (i.e., without context information; Reinisch et al., 2010) and in relation to preceding context. Moreover, it is plausible to assume that this mechanism has a prelexical locus. Studies on the effects of speaking rate on segmental distinctions suggest that rate effects occur early during processing (Miller & Dexter, 1988; Newman & Sawusch, 2009; Sawusch & Newman, 2000). When listeners rapidly judge which sound they hear, they treat rate information from a following vowel mostly as fast (Miller & Dexter, 1988). That is, they seem to use the available rate information before they processed the complete stimulus. Importantly, this use of rate information appears to be mandatory (Miller & Dexter, 1988) – listeners seem unable to ignore rate information. In addition, speaking rate is taken into account before stream segregation occurs (Newman & Sawusch, 2009; Sawusch & Newman, 2000). If available, listeners use rate information from speakers other than the target speaker. This early and mandatory use of rate information suggests that the rate adjustment mechanism operates at a prelexical processing stage. We suggest further that this is a unitary mechansim, that is, one through which speaking rate information is used to modulate the interpretation of segmental and suprasegmental cues.

In summary, the present series of experiments demonstrated that speaking rate context affects the perception of durational cues to the lexical stress patterns of Dutch multisyllabic words. Lexical stress patterns on multisyllabic words are distributed over more than one syllable and speaking rate shifts the perception of these patterns by affecting

whatever syllable it is adjacent to. The word's initial syllable, however, is most important for listeners to recognize stress patterns. Listeners based their decisions on initial syllable durations even if the syllable's stress status could only be disambiguated by the preceding rate context. The duration of second syllables was used to a lesser extent. The effect of rate was similar for different stress patterns and therefore independent of the location of primary stress in the word. Speaking rate thus provides important context information in spoken word recognition.

Speaking rate from proximal and distal contexts is used during word segmentation

---

Chapter 5

Reinisch, E., Jesse, A., & McQueen, J. M. (under revision). Speaking rate from proximal and distal contexts is used during word segmentation. *Journal of Experimental Psychology: Human Perception and Performance.*

Abstract

A series of eye-tracking and categorization experiments investigated the use of speaking rate information in the segmentation of Dutch ambiguous word sequences. Word boundaries are cued in part by the duration of their juncture phonemes (e.g., [s] in 'eens (s)peer' "once (s)pear", and closure duration of [t] in 'nooit (t)rap' "never staircase/quick"). Since durational cues ought to be perceived relative to speaking rate, the rate of a preceding context sentence should affect the perception of such word boundaries. The present study indeed shows that following a fast context sentence the juncture phonemes were perceived as longer, and hence more often as target-initial, than when following a slow context. Furthermore, listeners used speaking rate in the evaluation of durational information during word recognition as soon as that information became available. Additional experiments established that listeners used not only rate information from a context proximal to the target but also from a more distal context. Stronger effects of distal contexts, however, were observed in categorization tasks, which reflect the result of the word recognition process, than in eye tracking, which measures word recognition over time. Whereas in categorization experiments the amount of rate context had the greatest effect on the use of speaking rate, in eye tracking its proximal location was the most important. These findings constrain accounts of how speaking rate modulates the interpretation of durational cues during word recognition by suggesting that rate estimates are used to continuously evaluate upcoming phonetic information during a prelexical phase of processing.

Introduction

Speech unfolds over time, so duration is an important cue in spoken word recognition. Duration can be a segmental cue to phoneme identity (e.g., phonemically long vs. short vowels, voice-onset time duration as a cue to consonantal voicing), or a suprasegmental cue, for example to a word's lexical stress pattern or to the location of word boundaries. Durational cues, however, are not perceived in an absolute fashion but rather relative to the rate at which an utterance is spoken. If an utterance is spoken fast, segment durations shorten (Crystal & House, 1982; 1988; Gay, 1978). Listeners take this into account and interpret durational cues relative to rate information in the surrounding context (see Miller, 1981, 1987 for overviews). The effect of rate has been investigated extensively in the domain of phoneme perception (e.g., Allen & Miller, 2001; Miller, 1981; 1987; Miller & Dexter, 1988; Miller & Liberman, 1979; Miller & Wayland, 1993; Wayland, Miller, & Volaitis, 1994). Recently, it has also been demonstrated that rate modulates perception of duration as a suprasegmental lexical-stress cue (Reinisch, Jesse, & McQueen, in press). The present series of experiments investigated speaking rate in relation to another type of durational cue: duration as fine phonetic detail used in word segmentation (e.g., Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). We ask if, when, and how speaking rate influences the interpretation of the duration of juncture phonemes, and hence where word boundaries are placed in the segmentation of continuous speech.

We first show that speaking rate indeed influences word segmentation, and then address two additional questions. When during the word recognition process do listeners take speaking rate information into account, and what is the relative contribution of the proximity and the amount of rate contexts in the use of rate information? These questions were addressed with a categorization task that reflects the overall outcome of the word recognition process, but also with an eye-tracking paradigm that taps directly into the word recognition process as it happens. Examining these questions in both tasks provides additional insight into the mechanisms of how rate information from a preceding context influences the perception of word boundaries.

The focus will be on lexically ambiguous sequences in continuous speech. For example in Dutch, 'een speer' ("one spear") and 'eens peer' ("once pear") contain the same sequence of sounds but differ in which word the [s] belongs to. Word boundaries are marked by a variety of acoustic cues that listeners can use (e.g., Cho, McQueen, & Cox, 2007; Mattys, White, & Melhorn, 2005; Nakatani & Dukes, 1977) and duration is an important one of them (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, Liberman, Eccardt, &

Pesetsky, 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Burani, & Scott, 1995). The longer a boundary sound is (e.g., the [s] in the example above), the more it supports a word-initial interpretation (i.e., 'een speer').

Repp et al. (1978) showed that the interpretation of a boundary sound also depends on the durations of the following segments. The interpretation of a word sequence as 'great ship' vs. 'grey chip' depended on the relation between the closure duration of the [t] and the duration of the following [ʃ]. If the duration of the following fricative [ʃ] was long, the [t] was assigned to the previous word (i.e., 'great ship') more often than when the frication duration was short. That is, durations of the context following a word boundary can influence the perceived location of that boundary. In the present study, we first tested with a categorization task whether rate information from a sentence preceding the boundary sound can also influence its perceived location, as Repp et al. (1978) predicted. Following a fast rate context, a given boundary sound should be perceived as relatively long and thus more often as word initial. Following a slow rate context, the same sound should be perceived as shorter and thus more often as not word initial.

Moreover, since the use of durational cues to word segmentation has been shown to occur during the early phases of word recognition (Shatzman & McQueen, 2006) we also asked whether speaking rate information is taken into account at the same time or whether it is used later to reinterpret durational cues. Studies on the perception of speaking rate so far mostly used "offline" tasks such as phoneme categorization or category goodness judgments. These tasks measured the results of the phoneme or word recognition process by asking listeners to give explicit answers about what they perceived. Nevertheless, several studies on the effects of speaking rate suggest that rate effects may indeed occur early during processing. When, for example, listeners answered very rapidly in a categorization task, that is, when they initiated their answers before the complete context information could have been processed, then listeners tended to treat following rate context as fast, consistent with the amount of information they had processed at that time (Miller & Dexter, 1988).

Further support for the earliness of rate effects in phoneme perception comes from studies that demonstrated the use of rate information from talkers other than the target speaker on the interpretation of the target speaker's speech (Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000). Also, rate information of competing speech affects phoneme perception even if the competitor speaker is at a different spatial location from the target speaker (Newman & Sawusch, 2009). This suggests that rate information is taken into account at a prelexical stage of processing,

even before early processes such as perceptual grouping of speakers or stream segregation occur. None of these previous studies, however, directly assessed when during word recognition speaking rate is taken into account. To address this question our experiments used printed-word eye tracking to tap into the "online" word recognition process. This method allowed us to monitor over time how listeners' hypotheses about upcoming words are modulated by preceding speaking rate information.

Eye tracking is an excellent method to investigate the inner workings of the word recognition process. When listening to speech, listeners spontaneously fixate visual referents (including printed words; Huettig & McQueen, 2007; McQueen & Viebahn, 2007; Reinisch, Jesse, & McQueen, 2010) that match their current hypotheses about what is being said. Eye-tracking reveals what these hypotheses are and how they change over time (Allopenna, Magnuson, & Tanenhaus, 1998; Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995). Using eye tracking, it has been shown previously that listeners use duration as a cue to word boundaries as the speech signal unfolds (Shatzman & McQueen, 2006). When Dutch listeners heard segmentally ambiguous phrases of the type 'een(s) (s)peer' ("on(ce) (s)pear") they considered an s-initial word 'spuit' ("syringe") longer as a competitor of the target 'peer' ("pear") if the [s] was long than if it was short (Shatzman & McQueen, 2006). Note that in this study the duration of [s] was the only one of several acoustic measures that correlated with eye-movement behavior.

Here we used materials similar to those of Shatzman and McQueen (2006). The critical materials were Dutch segmentally ambiguous word sequences in which the first word always ended in one of two boundary sounds (i.e., [s] or [t]) and the second word did or did not start with the same sound (i.e., an [s] or [t], respectively). The question was thus whether the boundary sound was word-initial or not. Target words with the boundary sound [s] always followed the word 'eens' ("once"; s-trials, e.g., 'wel eens peer'). In this case listeners were explicitly told that 'eens' was the intended interpretation of the pre-boundary word as 'een' ("one") is also an existing Dutch word. Independent of the interpretation of the pre-boundary word, however, the longer the duration of the [s], the more likely the target word should be perceived as s-initial. Target words with the boundary sound [t] always followed the word 'nooit' ("never"; t-trials, e.g., 'nooit rap', "never quick"). 'Nooi' is not an existing Dutch word. Here again listeners had to decide whether the target word started with a [t]. In a sequence of two [t]s at a word boundary, the first one is commonly not released. (This was the case for our speaker.) Instead, the closure duration indicates the number of sounds. The longer the t-closure, the more likely it is that the first [t] has not been released and that the second word is t-initial.

Whereas for [s] the critical information is the duration of the frication noise, for [t] it is closure duration, that is, silence. The presence of a signal in [s] (i.e., frication noise) may be more informative than the absence of an audible signal during the t-closure. When hearing silence, the listener cannot be sure whether indeed silence was presented or whether some acoustic information has been missed. The duration of the frication in [s] may therefore be more easily interpreted relative to the context and lead to larger rate effects than the closure of [t], at least at an early stage of word processing. The use of the two boundary phonemes [s] and [t] thus allowed us to test whether speaking rate effects depend on the type of durational information that should be interpreted relative to rate.

Effects of speaking rate have not been shown with eye tracking before. Reinisch, Jesse, and McQueen (2008a) failed to detect such effects, and argued that this may have been because their stimuli were not entirely ambiguous. Absolute durational cues (in the case of Reinisch et al., 2008a, to the lexical stress patterns of words) may have disguised effects of the speaking rate context. The present series of experiments therefore attempted to maximize the chances of detecting an effect of rate by setting the critical boundary sounds ([s] and [t]) to durations which at a normal speaking rate would be ambiguous between initial and non-initial interpretations. The use of rate information should therefore help listeners recognize the targets. Relative to the rate-manipulated preceding fast or slow context the ambiguous boundary sounds should be interpreted as long or short.

The other major question of the present series of experiments concerned the relative roles of proximity and amount of rate context. For the interpretation of phoneme categories, preceding and following speaking rate information has been shown to exhibit a stronger effect the closer it is to the target phoneme. This research suggests that speaking rate is calculated within a small time window of about 250 to 300 ms from the target in both directions (Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). Rate context within this time window of approximately one syllable affects the perception of a target sound even if the rate-manipulated material is not adjacent (e.g., the neighboring segment) to the target sound (Newman & Sawusch, 1996; Sawusch & Newman, 2000). Here we investigated the relative effects of proximal and distal preceding context. Distal context preceding a target phoneme by more than 300 ms can affect target perception (Summerfield, 1981; Wayland et al., 1994). The effect of this long-range context on phoneme perception appears, however, to be different from the effect of the immediate 250-300 ms context (Wayland et al., 1994). The rate of utterance-length context affects only the location of the phoneme's best exemplar range but not the width of the best exemplar range. Nevertheless, rate information that is distal to the target can have an effect on target perception.

Previous research suggests that the relative role of distal and proximal rate information is modulated by the rhythmic pattern of an utterance (Kidd, 1989). In an utterance in which all stressed syllables were set to the opposite rate than the unstressed syllables, listeners interpreted the target phoneme relative to the rate of the stressed syllables, even though the stressed syllables were not adjacent to the target (Kidd, 1989). This effect of the non-adjacent stressed syllables was similar in size to the effects found in the condition where the complete utterance (i.e., stressed and unstressed syllables) was rate-manipulated. This suggests that the durations of stressed syllables are used to calculate rate. The rate of immediately adjacent unstressed syllables only affected perception when their rate was unexpected with regard to the preceding rate patterns (Kidd, 1989). The interaction between information from proximal and distal rate context may therefore depend on the rhythmic structure of the utterance.

In the present study, we further examined under which conditions distal rate context affects word segmentation by manipulating both the amount and relative location of the rate context. To assess the relative importance of distal vs. proximal context, we investigated effects of distal context when the proximal context was informative for the interpretation of the boundary sounds (i.e., proximal context was also rate-manipulated) and when it was not informative (i.e., proximal context had a speaking rate that did not disambiguate the target). This allowed us to test whether a distal rate context attenuates the effect of proximal context and whether a distal rate context can directly affect word recognition. Moreover, we asked whether a longer distal context would yield a stronger rate effect than a short distal context. By addressing these questions in both eye-tracking and categorization tasks, we can obtain insight into possible differences in the relative roles of location and amount of rate context in offline vs. online word recognition. Since during the word recognition process listeners continuously update their interpretation of speech, it seems plausible that they also continuously update their estimate of speaking rate from the incoming information. This need for continuous information update, however, may lead listeners to give most weight to the immediately preceding speaking rate information during word recognition. Therefore, in eye tracking, listeners may be influenced more by proximal than by distal rate context. In contrast, in categorization experiments listeners have more processing time and therefore have the opportunity to post-perceptually re-process speaking rate information from a longer context. Distal rate context in general and the amount of (distal) rate context in particular may thus play a larger role in determining categorization decisions than in influencing eye movements.

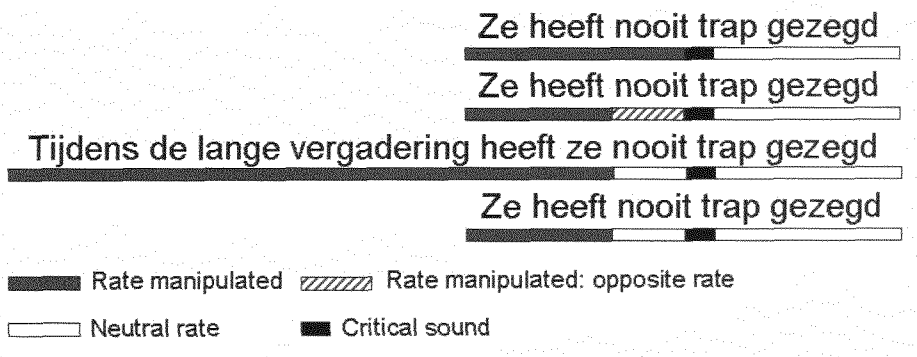

Figure 1. *Rate and context manipulations across experiments. Examples are given for one word of the t-trial word set. Manipulations for s-trials were identical (see text for details).*

In summary, the present series of experiments investigated the use of speaking rate information during the segmentation of words from a continuous speech signal. Following a fast rate context the perceived duration of ambiguous boundary sounds should be long, hence the target word should be temporarily interpreted as s- or t-initial. The opposite was expected for a slow rate context. Two different boundary sounds were examined to explore how rate context affects different types of durational cues (i.e., presence of a signal in [s] vs. the absence of a signal in the closure of [t]). Since listeners use duration as fine phonetic detail during the word recognition process, we asked whether speaking rate would be immediately taken into account for the interpretation of these cues. If this was the case, then we could assume a single mechanism for the interpretation of absolute and relative durational information in word recognition. Listeners would track rate information from the preceding context and use it to evaluate upcoming durational information optimally during a prelexical phase of processing. Subsequent experiments further explored the role of proximity and amount of rate context on word boundary perception. We asked whether distal rate context modulates the effects of context that is proximal to the target and whether distal rate context could be sufficient for listeners to disambiguate the boundary sounds if the proximal context was not informative about target-interpretation (see Figure 1 for an overview). In addition, we asked whether a longer distal context has more influence on target perception than a short distal context. The comparison of rate effects from proximal

and distal contexts during the word recognition process (eye tracking) to such rate effects when post-perceptual processing time is available (categorization) should inform us about the estimation and use of rate information in these different situations.

Experiment 1

Experiment 1 asked the basic question of this research project: Are durational cues to word boundaries interpreted relative to speaking rate of the preceding context? In a categorization task listeners had to decide whether the target word started with the boundary sound or not. As mentioned above, previous research has shown that the duration and hence the location of a juncture phoneme is interpreted relative to the duration of the following phoneme (Repp et al., 1978). The present experiment expands on these previous findings by examining the influence of an utterance-length preceding speaking rate context on the perceived location of the boundary sound. Two different boundary sounds were used: the duration of the frication noise in [s] and the duration of silence in the closure duration of [t].

The second purpose of Experiment 1 was to find perceptually ambiguous sounds for the subsequent eye-tracking experiments. That is, we wanted to find juncture sounds [s] and [t] which, at a normal speaking rate, are perceived equally often as word-initial sounds or not. This manipulation sought to maximize the chance of detecting the effect of rate online since, in a previous eye-tracking study on speaking rate effects (Reinisch et al., 2008a), absolute durational cues in the target words were sufficiently strong to mask possible context effects of rate. Using ambiguous boundary sounds in Experiment 2-5 should thus make effects of rate more easily observable as only the sounds' interpretation relative to the rate of the context can disambiguate their position.

In summary, Experiment 1 investigated whether speaking rate affects the perception of word boundaries, it tested whether the rate effect differs for the two types of boundary sounds we used, and it established the ambiguous durations of those sounds that were needed for the following eye-tracking experiments.

Methods

Participants

Twenty-four participants from the Max Planck Institute's participant pool took part for a small payment. They were all native Dutch speakers and reported no hearing problems.

Materials

Twelve participants categorized two word pairs from the s-trials of the subsequent eye-tracking experiments ('stillen'-'tillen', "satisfy"-"lift", and 'speer'-'peer', "spear"-"pear"), and twelve participants categorized two word pairs from the t-trial word set ('toost'-'oost', "toast"-"east", and 'trap'-'rap', "staircase"-"quick"). These words were a subset of the words recorded for the eye-tracking experiments (for a detailed description, see below). To create long endpoints for the [s] and [t] duration continua that are representative of targets starting with these boundary sounds (henceforth "initial" targets), the initial consonants of the chosen s- and t-initial items were replaced by tokens of [s] and [t] that had a duration of one standard deviation above the mean of all recorded s- or t-initial words (147 ms for [s], 138 ms for the [t]-closure). The critical durations (i.e., [s] and [t]-closure) were then manipulated along 15-step continua. Steps were created by successively cutting off approximately 8 ms (1/18 and 1/17 of the original duration for s- and t-trials respectively) from the end of the [s] or the t-closure until 33 ms of the [s] and 24 ms of the [t]-closure were left.

To minimize the influence of segmentation cues other than the critical durations, other potential cues were set to fixed values. The closure duration of stops following the [s] in s-trials was set to 80 ms, which was close to their mean duration in the s-initial words. The [t] in the t-trials was chosen such that the RMS-amplitude of the [t]-burst was within one standard deviation of the average RMS-amplitude in t-initial and not-t-initial words (i.e., 0.018 Pascal, mean: 0.023 Pascal, standard deviation: 0.007 Pascal). Carrier sentences were 'Ze heeft wel eens x gezegd' ("She once said x") for s-trials and 'Ze heeft nooit x gezegd' ("She never said x") for t-trials (for details see the description of eye-tracking materials below). The continua were spliced into the carrier sentences such that they replaced the last sounds in the sentences before the target words, that is, the [s] of 'eens' and the [t] of 'nooit'. The target words were attached without their initial [s] and [t] sounds (i.e., 'peer', 'tillen', 'rap', and 'oost' excised from recorded 'speer', 'stillen, 'trap', and 'toost'). All splicing was done at

positive-going zero crossings. The parts of the carrier sentences that preceded the critical sounds were rate-manipulated (see Figure 1). Carriers were presented at a fast (66% of original), normal (original), or slow (133% of original) rate. The rate-manipulation was implemented linearly using the PSOLA algorithm provided in Praat (Boersma & Weenink, 2007).

Procedure

For both s- and t-trials each rate context was combined with each word at each step of the continua. Each combination was repeated six times. This resulted in 540 trials per participant (2 Word pairs x 3 Rates x 15 Steps x 6 Repetitions). The experiment was blocked by word pair with order of pairs counterbalanced across participants. Presentation order within blocks was randomized for every participant with the restriction that all combinations of step and rate would be presented once before a repetition occurred. Speaking rates were intermixed at random.

Throughout each trial participants saw the two response alternatives on the screen with the s- or t-initial word on their left (e.g., speer peer, trap rap). Auditory stimuli were presented over headphones at a comfortable listening level, 200 ms after the words appeared on the screen. Listeners had to indicate by button press as quickly and as accurately as possible which of the two words they heard in the sentence. If the response time exceeded 3000 ms from the onset of the critical sounds the next trial started automatically. Every 45 trials participants were given a short break. The experiment was controlled by NESU experimental software (http://www.mpi.nl/world/tg/experiments/nesu.html).

Analysis

Results were analyzed separately for s-trials and t-trials using linear mixed effect models (Baayen, Davidson, & Bates, 2008) with a logit link function to deal with the dichotomous dependent variable (i.e., initial response (= 1) vs. non-initial (= 0) response). Rate (-.5 = fast, 0 = normal, .5 = slow), step (0-14), and their interaction were entered in the model as fixed factors, and participant and item as random factors. The model maps a factor's condition with the value 0 on the intercept and assigns regression weights for the adjustments required to map from this condition to every other level of the factor. In this and all subsequent experiments, non-significant interactions were eliminated and the models were refitted. Only significant results will be reported.

Results

Figure 2 shows categorization data in response to the minimal pairs from the s-trials and t-trials.

S-trials

For s-trials a main effect of step as well as an interaction between rate and step were found ($b_{step} = 0.33$, $p < .001$; $b_{step*rate} = -0.07$, $p < .001$). More s-initial responses were given the longer the s-duration and the effect of rate was stronger the longer s-durations were. To find the perceptually most ambiguous duration at which a rate effect was present, models with rate as fixed factor and participant and word as random factors were fitted for every step of the continuum. Since normal rate (value 0) was mapped onto the intercept, a non-significant intercept can be taken as indication of ambiguity, that is, that there was no preference for either initial or non-initial responses for the respective step in a normal rate context. This was the case for steps 5 to 7 (step 5: $b_{intercept} = -0.17$, $= .46$; $b_{rate} = -0.11$, $p = .66$; step 6: $b_{intercept} = -0.05$, $p = .78$; $b_{rate} = -0.69$, $p < .01$; step 7: $b_{intercept} = 0.54$, $p = .07$; $b_{rate} = -0.95$, $p < .001$). A duration between step 6 and 7 (i.e., 86 ms) was selected to be the most suitable ambiguous sound as steps 6 and 7 showed the largest numerical difference in initial responses between fast and slow rate, and a rate effect was present at these steps for 9 out of 12 participants. The vertical line in the left panel of Figure 2 shows the location of this duration on the s-continuum.

T-trials

For t-trials overall main effects of rate and step were found ($b_{rate} = -1.29$, $p < .001$; $b_{step} = 0.55$, $p < .001$). More t-initial responses were given the longer the t-closure and the faster the rate of the context. Separate analyses for each step of the continuum suggested that step 5 was perceived as ambiguous at a normal rate context ($b_{intercept} = 0.02$, $p = .93$; $b_{rate} = -1.79$, $p < .001$). An intermediate duration between step 5 and 6 (i.e., 69 ms) was chosen as the ambiguous sound for t-trials in the eye-tracking experiments. At both steps a large difference between responses in the fast vs. the slow rate context was found and this rate effect was present for a large number of participants at these steps (all participants showed an effect at step 5, and 11 out of 12 participants at step 6). The vertical line in the right panel of Figure 2 shows the location of this duration on the t-continuum.

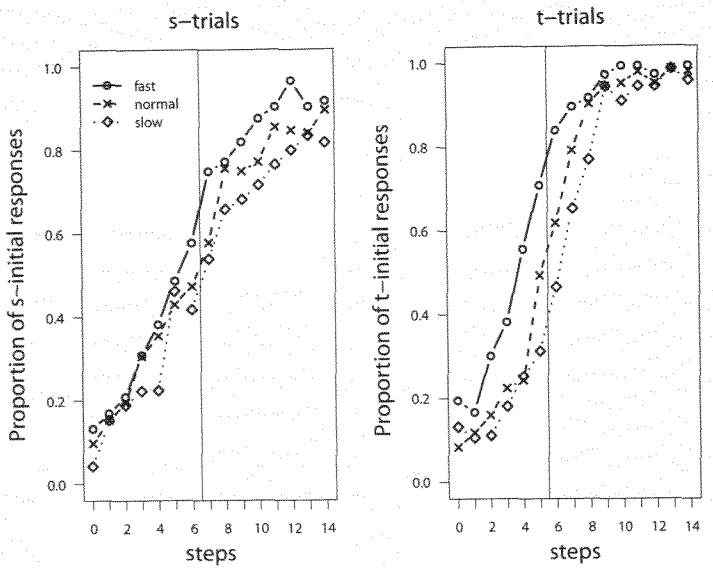

Figure 2. *Proportion of initial responses for s-trials and t-trials in Experiment 1. The duration of [s] and the duration of the t-closure were varied along 15-step continua. Preceding sentence context was presented at a fast, normal, or slow speaking rate. The vertical line in each panel marks the duration that was chosen for the eye-tracking experiments.*

Discussion

Listeners use speaking rate information in the identification of word boundaries that are cued by durational information. The faster the rate context, the longer the boundary sound was perceived, and thus the more likely it was interpreted also as word-initial. This result held for both types of juncture phonemes. Speaking rate affected the perceived duration of the frication of [s] as well as the duration of silence during the closure of [t]. Note that previous results on the relation between speaking rate context and the perception of word-initial stop closure durations were mixed. Listeners in some studies interpreted closure duration as part of the stop and thus as a cue to voicing rather than as a cue to rate information (Newman & Sawusch, 2009; Summerfield, 1981; Port, 1979). In at least one study, however, closure duration was used as cue to speaking rate information (Kidd, 1989). Note, however, that these previous studies tested sounds in English. Here, we used Dutch word sequences. In contrast to English where voicing is cued mainly by closure duration

(Klatt, 1976), the main cue to voicing in Dutch is the presence of vocal vibration during the closure (van Alphen & Smits, 2004). Here we showed that the duration of the [t]-closure was interpreted relative to speaking rate and in this way could also be used as a cue to word boundaries in Dutch.

Experiment 1 replicated previous findings on the use of duration as cue to word boundaries, and findings that durational information is interpreted relative to the speaking rate of the preceding context. By combining these two research questions, however, Experiment 1 extended this research. It showed that the perception of word boundaries cued by duration depends on the speaking rate of the preceding context. Experiment 2 adds a third line of previous research to our investigation: the question about when during the word recognition process speaking rate is taken into account. Experiment 1 established durations at which the boundary sounds were ambiguous for their interpretation at a normal speaking rate but showed a large rate effect for a large number of participants. These durations were used in all the following eye-tracking experiments.

Experiment 2

Experiment 2 introduced an eye-tracking paradigm to investigate whether and when speaking rate information is used to interpret durational cues online during the word recognition process. Since listeners use duration as fine phonetic detail to segment words as the speech signal unfolds, speaking rate should modulate the interpretation of these durational cues. If listeners interpret the ambiguous critical sounds relative to speaking rate and use these relative durations in the same way they used the sounds' absolute durations in an earlier study (Shatzman & McQueen, 2006) we expect a comparable time course of rate effects here as was found previously for unambiguous long or short boundary sounds. As in Experiment 1, in Experiment 2 the complete context preceding the boundary sounds was rate-manipulated. The demonstration of online rate effects would suggest that listeners take into account not only local durational information but that acoustic information from a larger stretch of speech stays available for the interpretation of upcoming acoustic information. Such a result would thus confirm previous suggestions that speaking rate is used at an early stage of word or phoneme processing. In addition, the use of eye tracking allowed us to examine the time-course of the speaking rate effect.

A categorization task with the same rate manipulation as in eye tracking was added to the experiment to compare participants' rate effects in online and offline word

recognition. We asked whether the use of rate information during word recognition would be correlated with effects that are measured in a task tapping into the result of this recognition process.

Methods

Participants

Thirty Dutch native speakers from the Max Planck Institute's participant pool were paid for taking part. They all had normal hearing and normal or corrected-to-normal vision. None had taken part in Experiment 1.

Materials

Eye tracking

Fifty-four Dutch words of one or two syllable length were chosen as targets for s-trials and another fifty-four words for t-trials. Eighteen target words from each of these sets were s- or t-initial. The remaining thirty-six target words in each set were not-s-initial or not-t-initial ("non-initial") but chosen such that the addition of an [s] or [t] at the beginning of the word would result in another Dutch word, for example, 'peer' and 'rap', where 'speer' and 'trap' are also Dutch words (see Appendix). A subset of these minimal pairs was used in Experiment 1 and in the categorization task described below. For the eye-tracking task, displays of four printed words were created for each trial where each target was presented along with a phonological competitor and two distractors. Competitors overlapped segmentally with the target up to and including the first vowel (see Appendix). For initial targets the competitors were non-initial (e.g., 'peper' and 'oester' - "pepper" and "oyster" - for the targets 'speler' and 'toetsen' - "player" and "test") and for non-initial targets the competitors started with an [s] or [t] (e.g., 'speen' and 'tractor' - "pacifier" and "tractor" - for the targets 'peer' and 'rap'). Words from a minimal pair never occurred together on the screen. Target and competitor sets were matched on their average CELEX lexical frequency (Baayen, Piepenbrock, & Gulikers, 1995; s-initial: $t(17) = 0.01$, $p = .99$; t-initial: $t(17) = -0.31$, $p = .76$; not-s-initial: $t(35) = -0.75$, $p = .49$; not-t-initial: $t(35) = 1.58$, $p = .12$). Each target-competitor pair was assigned two phonetically, orthographically, and semantically unrelated distractors. Thirty-six filler and six practice displays, each containing four unrelated words,

were constructed such that the targets did not begin with [s] or [t] nor formed a minimal pair by adding an [s] or [t] at the beginning.

A female native Dutch speaker recorded multiple tokens of the targets (x) as well as of their minimal pairs (if available) in the carrier sentences 'Ze heeft wel eens x gezegd' for s-trial items or 'Ze heeft nooit x gezegd' for t-trial items spoken at a normal rate. Sentence accent was placed on the target word. One token of each carrier preceding the targets and two versions of 'gezegd' were selected as the contexts into which all targets were spliced. A token of 'gezegd' was assigned independently of trial type so that it maximized acoustic coherence between the target and its following context. The preceding carriers were chosen such that the selected sentence was of average duration compared to all recorded carriers. Using the same sentence for all items in the respective trial set controlled for variation in pitch contours. The speaking rate of the carrier sentences was linearly compressed or expanded using PSOLA as provided in Praat (Boersma & Weenink, 2007). The whole preceding context up to the ambiguous sounds was rate-manipulated (see Figure 1). The carrier sentences excluding their final sounds (i.e., the [s] and [t] in 'eens' and 'nooit' respectively) were speeded up to 66% of their original duration for the fast speaking rate condition and were slowed down to 133% of their original duration for the slow speaking rate condition. The targets and the part of the carrier sentence that followed the target were presented at a normal rate. Note that in contrast to Experiment 1 the preceding context was never presented at a normal speaking rate.

The best token of each target word was selected using two criteria. First, to control for speaking rate, the sentence it had been recorded in matched as closely as possible the duration of the carrier sentence that was selected for manipulation. Second, the target token had to fit the selected context in terms of pitch and voice characteristics to minimize audibility of the splicing manipulation. To make the words perceptually ambiguous between initial and non-initial interpretations, targets for the non-initial conditions were created from their s- or t-initial minimal pairs. Vowel-initial words were occasionally produced with creaky voice or delayed voicing relative to the offset of frication. Using these tokens as base for the stimuli could have produced a bias towards non-initial targets.

Critically, the same token of [s] and [t] was used for all items in initial and non-initial conditions as well as for all filler items in the respective trials. To maximize the use of speaking rate in recognizing the target word, the critical sounds were set to their perceptually ambiguous durations, as established in Experiment 1. The ambiguous durations were 86 ms for the [s], and 69 ms for the closure of the [t]. At a normal rate context

the interpretation of these target durations was equally likely to be initial or non-initial (see Figure 2). The stimuli in the eye-tracking experiment thus consisted of a concatenation of the rate-manipulated preceding carrier sentences excluding the final [s] or [t] in 'eens' and 'nooit', the ambiguous sounds with the durations established in Experiment 1, the target words without the initial [s] or [t], and the following carrier ('gezegd').

Categorization

One word of each trial set (s-trials: 'peer', t-trials: 'rap') was selected as target. The words came from the non-initial target condition of the eye-tracking experiment and formed a minimal pair if an [s] or [t] was added at their beginnings (i.e., 'speer' and 'trap'). These word pairs were also used in Experiment 1.

Whereas in Experiment 1 the whole duration continuum of the critical sounds was presented, a subset of steps was selected for the categorization experiment here. The chosen steps were not equally spread over the original duration continuum but rather contained a higher density of steps in the ambiguous region. For s-trials the steps used for categorization had durations of 33, 57, 74, 82, 90, 98, 115, and 147 ms (steps 0, 3, 5, 6, 7, 8, 10, and 14). For t-trials these were 24, 41, 57, 65, 73, 81, 105, and 138 ms (steps 0, 2, 4, 5, 6, 7, 10, and 14). The same rate-manipulated sentences as in the eye-tracking experiment were used in the categorization task. That is, the part of the carrier sentence that preceded the critical sounds was speeded up to 66% or slowed down to 133% of its original duration. The target words and carriers that followed the targets (i.e., 'gezegd') were presented at a normal rate. Unlike in Experiment 1 the part of the carrier sentence that preceded the target was never presented at a normal rate.

Procedure

Eye tracking

Listeners were positioned 65 cm in front of a 40.5 by 30.5 cm screen. Eye-movements were recorded with an SR Research Eyelink 1000 system at a sampling rate of 1000 Hz and controlled by Experiment Builder software (SR Research). Each participant heard half of the sentences of each trial type (s-trials and t-trials) in the fast rate context and half of the sentences in the slow rate context. Trial types and rate contexts were intermixed and presented in a different random order to each participant. The assignment of words to the four positions on the screen was randomized for each participant with the number of times

the target appeared on each position balanced within each condition. Before the start of the experimental list all participants responded to the same six practice trials. Half of the practice items were s-trials, the other half were t-trials.

On every trial participants saw a fixation cross for 500 ms in the middle of the screen. Immediately afterwards four printed words appeared in the four quadrants of the screen. The center of each word was placed 13.5 cm from the nearest edge of the screen on the horizontal axis and 10 cm on the vertical axis. Words were presented in monospaced, lower case, Lucida Console font, size 20. The auditory stimuli were presented over headphones at a comfortable listening level. The acoustic onset of the sentences was timed such that 1800 ms after display onset listeners heard the critical sound (i.e., [s] or [t]). Listeners were instructed to click with the computer mouse on the word they heard in the sentence. The next trial started 500 ms after the listeners responded. Every 10th trial a drift correction was carried out to adjust for possible head movements. The eye-tracking experiment lasted approximately 15 minutes and was immediately followed by the categorization task.

Categorization

Each step-rate combination was presented 10 times. The order of trials was randomized for each participant with the restriction that all step-rate combinations were presented before a repetition occurred. Rates were mixed at random. The experiment was blocked by trial type, switching to the word pair of the other type after five repetitions of all steps and rates. Once within each block as well as between blocks participants were given a short break. Order of block was balanced across participants.

On each trial participants saw the words of the minimal pair on the screen, with the s- or t-initial word always shown on their left. After 600 ms the auditory stimulus was presented over headphones at a comfortable listening level. Participants were instructed to indicate which of the words on the screen they heard in the sentence by pressing one of two marked keyboard buttons. The words stayed on the screen throughout the trial until 500 ms after the participant responded or 3000 ms after the onset of the critical sound. 300 ms after either event the following trial was initiated. The categorization tasks were controlled by Experiment Builder software (SR Research).

Analyses

Eye-tracking

Eye-movement data were analyzed in time bins of 4 ms. Only data which were classified by the eye-tracking system as being fixations were taken into account. A fixation was counted as falling on a word if it was located within a predefined circle of diameter 11 cm around the center of the word. Only trials during which all registered positions of fixations fell on the screen, and only trials on which participants clicked on the correct word were analyzed. Further, trials containing the not-s-initial target 'tillen' and the t-initial target 'triest' were excluded from all analyses. Their competitors differed in vowel length and thus did not show the same amount of phonemic overlap with their targets as the other target-competitor pairs.

The time windows for analyses were chosen from 286 ms to 1000 ms after the onset of the critical sounds for s-trials and from 269 ms to 1000 ms for t-trials. The start of the time windows was defined by the offset of the critical sounds shifted by 200 ms. A time lag of 200 ms is commonly used as an estimate of the time needed to program and launch a saccade (see, e.g., Hallett, 1986; Matin, Shao, & Boff, 1993). That is, on average 200 ms after the acoustic input eye movements start reflecting the impact of this acoustic input on the word recognition process. Note that the durational information of the critical sounds could have been processed at the earliest only at critical sound offset and thus could potentially only show an effect at that time or later.

Eye-movements were analyzed separately for s-trials and t-trials in initial and non-initial conditions using linear mixed effect models (Baayen et al., 2008) as provided in the lme4 package (Bates & Sarkar, 2007) in R (version 2.8.0; The R foundation for statistical computing). The dependent variable was the difference between logistically transformed fixation proportions to targets and competitors (see Barr, 2008). Participant and item were entered as random factors. Speaking rate was entered in the model as numerical fixed factor (-.5 = fast; .5 = slow). The model maps the hypothetical value of rate = 0 on the intercept and estimates a regression weight with which the values of the factor have to be multiplied to move from the intercept to the respective factor levels. A positive regression weight indicates a greater difference between fixations on target and competitor following a slow than a fast speaking rate context. A negative regression weight indicates a greater difference between fixations on target and competitor following a fast than a slow speaking rate context. P-values were based on Markov chain Monte Carlo (MCMC) sampling.

Categorization

The categorization experiment was analyzed in a similar fashion as Experiment 1 by using linear mixed-effects models with a logit link function. The dependent variable was initial (= 1) vs. non-initial (= 0) responses. Participant was entered as random factor. Rate (fast = -.5, slow = .5), step, and their interaction were fixed factors. Since in Experiment 2 the durations of the continua were not equally spaced, durations of the chosen steps in milliseconds rather than their step numbers were used as numeric values for this factor. In addition, these new values of step were centered on the duration that was used for the eye-tracking stimuli (i.e., 86 ms for s-trials, and 69 ms for t-trials). That is, after centering, this duration had the value 0. In this way the main effect of rate referred to the duration of the critical sounds used in eye tracking.

Results

Eye-tracking

Nine trials (0.02%) had to be excluded due to fixations outside the screen. Another 102 trials (2.07%) were excluded because listeners clicked on either the competitor word (51 trials; 1.13%) or outside the defined fixation areas around the words (42 trials; 0.93%).

Figure 3 shows fixation proportions on target, competitor, and the average of the two distractors over time plotted for each rate from the onset of the critical sounds for s- and t-trials in the respective initial and non-initial conditions. The outer two vertical lines mark the time window of analysis. The left inner vertical line marks the end of the last segment that is shared by target and competitor ("target-competitor divergence point") shifted by 200 ms. The right inner line marks the target offset shifted by 200 ms. Both inner lines represent average values for each condition (see Table 1).

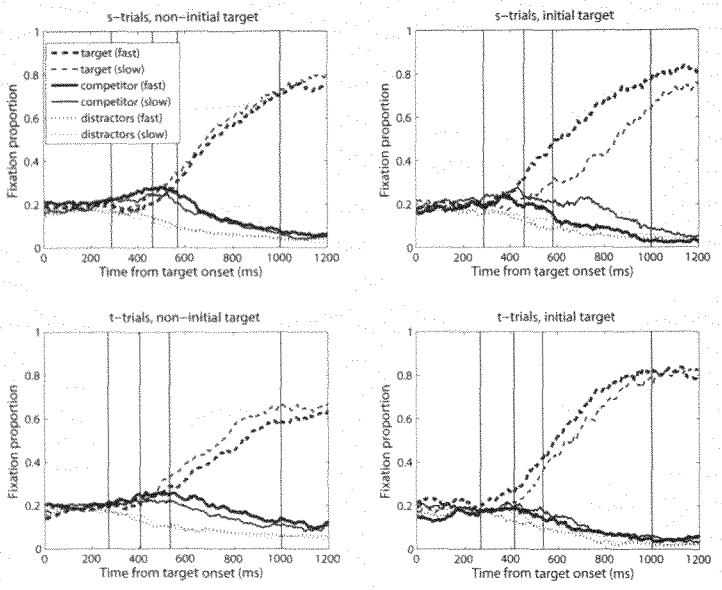

Figure 3. *Fixation proportions over time in the eye-tracking task in Experiment 2 to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines) following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. The outer two vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1000 ms after critical-sound onset. The inner two vertical lines mark the target-competitor divergence point and the end of the targets shifted by 200 ms averaged over condition (see text for details).*

	s-trials		t-trials	
	initial	non-initial	initial	non-initial
Target-competitor divergence point	263	259	205	213
Word duration	369	382	332	336

Table 1. *Durations measured from the onset of the critical sounds [s] and [t] until the end of the last segment that was shared by target and competitor (target-competitor divergence point) averaged per condition, and average duration of target words including the critical sounds. Durations are given in milliseconds.*

		s-trials				t-trials			
		s-initial		not-s-initial		t-initial		not-t-initial	
	n	*b*	p_{lmem}	*b*	p_{lmem}	*b*	p_{lmem}	*b*	p_{lmem}
Experiment 2	30	-2.19	< .001	0.39	= .147	-0.76	< .05	0.94	< .001
Experiment 3	26	-0.31	= .474	-0.43	= .170	-0.53	= .196	0.31	= .337
Experiment 4	30	0.61	< .05	-0.74	< .05	-0.13	= .738	0.43	= .107
Experiment 5	30	-0.63	= .117	0.03	= .927	-0.05	= .871	0.02	= .949

Table 2. *Results of eye-tracking experiments. Effects of speaking rate on the difference of log-transformed fixation proportions between target and competitor for s-trials and for t-trials. The time window spans the time from critical sound-offset shifted by 200 ms until 1000 ms after critical sound-onset. n = number of participants.*

Results of Experiment 2 are summarized in Table 2. For the s-trials an effect of speaking rate was found for s-initial targets (e.g., 'speler'). Following a fast rate context the [s] sounded relatively long and in this way supported the interpretation of the target as being s-initial. In contrast, following a slow rate context the critical [s] sounded relatively short and thus temporarily supported the not-s-initial competitor (e.g., 'peper') more than the s-initial target ('speler'). For not-s-initial targets (e.g. 'peer') a trend towards a rate effect in the opposite direction from that for the s-initial targets can be seen in the graphs. This effect was in the predicted direction, but failed to reach significance. For the t-trials an effect of speaking rate was found in the initial as well as in the non-initial condition. Following a fast rate context the closure duration of the [t] sounded longer, which led listeners to assume a word boundary during the closure. The tendency to fixate t-initial targets (e.g., 'toetsen') more than not-t-initial competitors (e.g., 'oester') was thus stronger following a fast than a

slow rate. As expected, the reverse pattern was found for not-t-initial targets (e.g., a stronger tendency to look at 'rap' than at 'tractor' in slow rate contexts).

Categorization

Figure 4 shows the categorization data in response to the minimal pairs along the duration continuum following a fast or a slow rate context. Table 3 summarizes the results. For s-trials effects of rate, step, and their interaction were significant. Listeners gave more s-initial responses the longer the duration of the [s] and more s-initial responses were given following a fast than a slow rate context. The interaction indicates that the effect of speaking rate was stronger the longer the s-durations. For t-trials only main effects of rate and step were found. Listeners gave more t-initial responses the longer the closure duration and following a fast than a slow rate context.

Additional analyses showed that the size of participants' rate effects in offline categorization did not predict their use of rate information in eye tracking.

		s-trials		t-trials	
	Factors	*b*	*p*	*b*	*p*
Experiment 2	rate	-1.39	< .001	-1.15	< .001
	step	0.07	< .001	0.11	< .001
	rate*step	-0.01	< .01	--	--
Experiment 3	rate	-0.43	< .001	-0.37	< .001
	step	0.05	< .001	0.10	< .001
	rate*step	-0.01	< .001	--	--
Experiment 4	rate	-1.19	< .001	-1.05	< .001
	step	0.07	< .001	0.12	< .001
	rate*step	--	--	-0.04	< .001
Experiment 5	rate	-0.26	< .001	-0.19	= .07
	step	0.06	< .001	0.11	< .001
	rate*step	--	--	-0.02	< .05

Table 3. *Effects of speaking rate, duration of the critical sounds (step), and their interaction in the categorization experiments. Non-significant interactions that were eliminated from the model are marked with "--".*

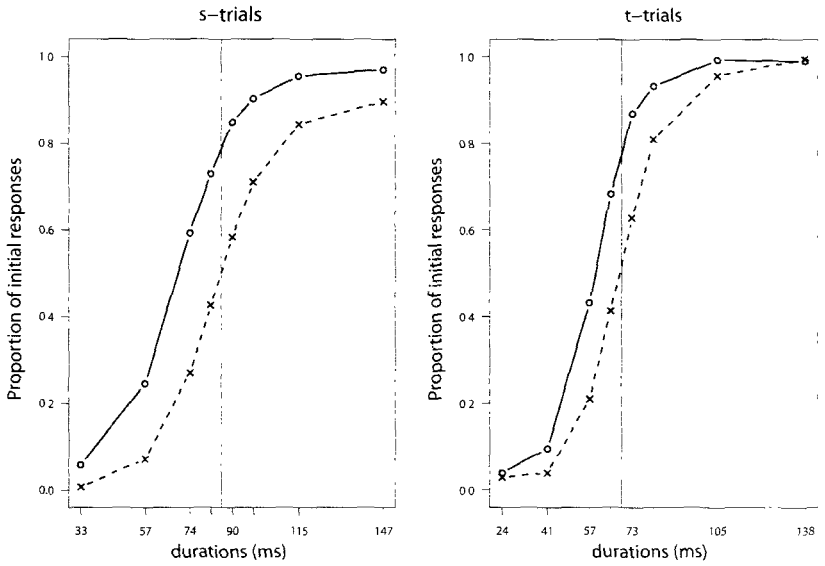

Figure 4. *Proportion of initial responses for s-trials and t-trials in the categorization task in Experiment 2. Duration of [s] and closure duration of [t] were varied on a subset of 8 steps of the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment.*

Discussion

Listeners use speaking rate information to disambiguate word sequences. The categorization results confirm the finding from Experiment 1 that speaking rate affects the perception of durational cues to word boundaries offline. Critically, the eye-tracking results show for the first time that speaking rate is used for lexical disambiguation during the word recognition process. S- or t-initial words were considered as targets more following a fast than a slow rate context. The opposite was the case for the non-initial targets. This result held for both types of juncture phonemes. Speaking rate affected the perceived duration of the frication of [s] in a similar fashion to the way it affected the perceived duration of silence during the closure of [t].

Overall the time-course of rate effects was very similar to the time-course of effects of boundary sounds that provided absolute durational cues (Shatzman & McQueen, 2006). As can be seen from Figure 3, speaking rate is used to modulate lexical competition over an

extended amount of time. Rate information started to affect word recognition while the acoustic signal of the target was unfolding. The effect, however, extended beyond the end of the words. That is, speaking rate information modulated speech processing even after the targets had been heard in their entirety. Note that this long-lasting target preference based on the rate context is what is likely to be reflected in categorization tasks where listeners respond after target offset (in our task on average 800 ms after the onset of the critical sound). Differences in the time course of competition between initial and non-initial target conditions reflect the delayed support of non-initial targets relative to their initial counterparts. Whereas initial words received support from the onset of the critical sound onwards, non-initial words received support only once the critical sounds had been processed.

In summary, Experiment 2 demonstrated that speaking rate is used during word segmentation. Ambiguous sounds are interpreted in relation to the preceding context and thus can be used to modulate the word recognition process. Rate effects were found for both boundary sounds. The following experiments further explored the conditions under which speaking rate context is used. In the next experiment, we investigated the role of proximity of rate context to the boundary sounds.

Experiment 3

Experiment 3 asked whether the effect of speaking rate in Experiment 2 was based on all preceding rate context or rather only the immediately adjacent context, as has been suggested in earlier studies (e.g., Summerfield, 1981). To show an influence of distal rate context, we investigated whether a distal rate context that had the opposite rate from the proximal context would attenuate the effect of the proximal context. The proximal context was defined as the words directly adjacent to the target (i.e., 'wel eens' for s-trials and 'nooit' for t-trials; see Figure 1). The distal contexts were the remaining initial parts of the carrier sentences used in Experiment 2 (i.e., 'ze heeft' for both trial types). Although the proximal context differed in the number of syllables between s-trials and t-trials, its absolute duration was equal across trial types (i.e., 112 ms in the fast condition, and 250 ms in the slow condition).

There are two accounts of how distal rate context could influence effects of the proximal context. An attenuation of the effect of proximal context relative to Experiment 2 would support accounts that argue for the use of rate information averaged over a time

window of approximately 300 ms around the target sound (e.g., Newman & Sawusch, 1996; Sawusch & Newman, 2000). Note that some of the distal context in our experiment would fall into this time window even for the slow rate proximal context (which had a duration of 250 ms) and therefore allowed for a potential attenuation of the effect of proximal rate by distal rate. In contrast, an enhancement of the effect of proximal context relative to Experiment 2 would be predicted if the rate of the proximal context was perceived as unexpected relative to the distal context (Kidd, 1989). That is, the distal context (e.g., slow) should make the opposite rate of the immediately preceding context sound even more prominent (e.g., faster) and thus lead to a greater effect on the perception of the critical sounds than if no rate-change had occurred (i.e., as in Experiment 2).

The use of eye tracking and categorization will again allow for a comparison between rate effects in online and offline processing. Distal rate context may have less influence during the word recognition process than offline. In the online situation listeners may rely on a continuously updated estimate of rate. If so, then that estimate presumably gives more weight to the immediately preceding rate context than to more distal contexts. In categorization, where listeners have time for additional post-perceptual processing, rate from distal context could be re-processed and thus play a larger role than online.

Methods

Participants

26 participants from the same population as in the previous experiments were paid for their services. None of the participants had taken part in the previous experiments.

Materials, Design, Procedure

The auditory stimuli of Experiment 2 were manipulated such that the initial portions of the sentences (i.e., 'Ze heeft') had the opposite rate than the portion immediately preceding the critical sounds (i.e., 'wel eens' and 'nooit', respectively). This was done by exchanging the distal contexts of the fast and slow carrier sentences from Experiment 2. All other aspects of the eye tracking as well as categorization tasks including their analyses were identical to Experiment 2. Fast and slow rate were coded as referring to the immediate context.

Results

Eye tracking

Ten trials had to be excluded due to fixations outside the screen (0.26%), and 79 trials (2.03%) for participants clicking outside the target area (53 of these clicks were on the competitor). Again all trials containing the targets 'tillen' and 'triest' were excluded for insufficient phonemic overlap with their competitors. Figure 5 plots fixation proportions to targets, competitors, and the average of the two distractors over time from the onset of the critical sounds. The graphs suggest that despite a mismatching distal context listeners used the speaking rate information of the immediately preceding rate context to recognize the targets in the initial conditions of both trial sets. Statistical analyses, however, indicate that this numerical effect of speaking rate was not significant (see Table 2). For s-trials and for t-trials there was no difference between speaking rate conditions in the recognition of either initial or non-initial targets.

Categorization

Figure 6 shows categorization data in response to the minimal pairs along the duration continuum in Experiment 3. As is evident from Table 3, main effects of rate and step were found for s-trials and for t-trials. More initial responses were given following a fast than a slow immediate rate context and more initial responses were given the longer the critical sounds. Moreover, as indicated by the interaction between step and rate, for s-trials the effect of rate was stronger the longer durations of the continuum.

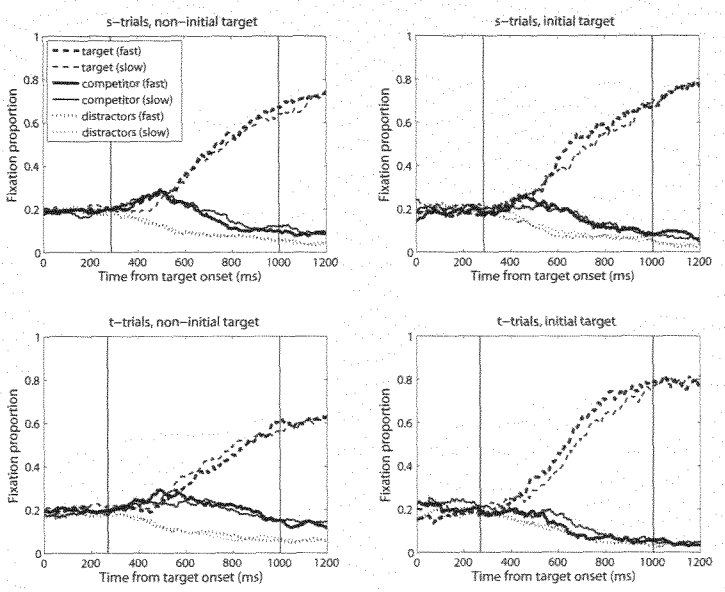

Figure 5. *Fixation proportions over time in the eye-tracking task in Experiment 3 to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines) following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. Fast and slow rate are coded for the rate of the immediate context. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1000 ms after the onset of the critical sounds.*

Cross-experiment comparisons

To investigate the influence of distal rate context on the use of context immediately preceding the target, eye-tracking and categorization data from Experiment 3 were compared to Experiment 2. For cross-experiment comparisons, the additional predictor experiment (levels: Experiment 2, Experiment 3) and its interactions with the other factors were entered into the statistical models. Non-significant interactions were eliminated and the models were refitted. A significant interaction between rate and experiment would indicate that the use of immediate context in Experiment 3 was modulated by information extracted from the distal context.

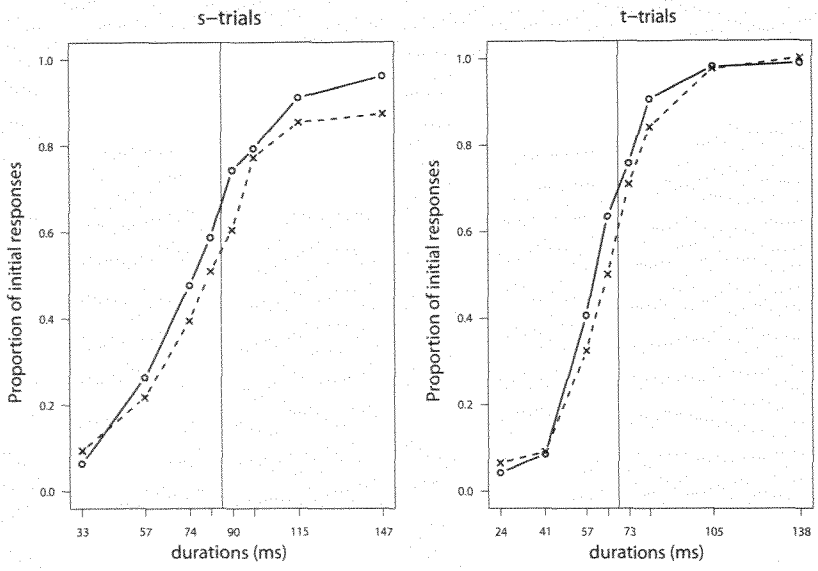

Figure 6. *Proportion of initial responses for s-trials and t-trials in the categorization task in Experiment 3. Duration of [s] and closure duration of [t] were varied on a subset of 8 steps from the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment. Fast and slow rate are coded for the rate of the immediate context.*

Eye tracking

Table 4 summarizes the results for the cross-experiment comparisons of the eye-tracking data. For s-trials a significant interaction between rate and experiment was found for s-initial and not-s-initial trials. The effect of rate was stronger in Experiment 2 than Experiment 3, suggesting that distal context modulates word recognition. In addition, main effects of rate were found for both conditions of s-trials. For t-trials the interaction between rate and experiment was not significant for t-initial and not-t-initial targets. In simpler models only containing main effects, an immediate rate context effect was found for t-trials.

		s-trials				t-trials			
		s-initial		not-s-initial		t-initial		not-t-initial	
	Factors	*b*	$p_{(mcmc)}$	*b*	$p_{(mcmc)}$	*b*	$p_{(mcmc)}$	*b*	$p_{(mcmc)}$
Exp.3 vs. Exp.2	rate	-4.06	< .001	1.21	= .055	-0.64	< .05	0.64	< .005
	experiment	-0.07	= .832	-0.27	= .380	-0.34	= .238	-0.44	= .105
	rate*experiment	1.87	< .001	-.082	< .05	--	--	--	--
Exp.4 vs. Exp.2	rate	-2.90	< .001	0.51	< .01	-0.44	= .112	0.67	< .001
	experiment	-0.13	= .410	0.08	= .610	-0.15	= .298	0.05	= .749
	rate*experiment	0.72	< .01	--	--	--	--	--	--
Exp.5 vs. Exp.4	rate	-0.68	< .05	0.32	= .101	-0.09	= .737	0.21	= .250
	experiment	0.06	= .837	-0.13	= .663	-0.27	= .344	-0.08	= .790
	rate*experiment	--	--	--	--	--	--	--	--
Exp.5 vs. Exp.3	rate	-0.19	= .520	0.22	= .313	0.22	= .424	-0.14	= .490
	experiment	-0.06	= .684	0.14	= .353	-0.11	= .438	0.23	= .086
	rate*experiment	--	--	--	--	--	--	--	--

Table 4. *Cross-experiment comparisons of the rate effect in the eye-tracking experiments for s-trials and t-trials with initial and not-initial targets. Non-significant interactions that were eliminated from the model are marked with "--".*

Categorization

Results for the cross-experiment comparisons of the categorization task are summarized in Table 5. For both trial types main effects of rate and step were found, as well as an interaction between these two factors. More initial responses were given following a fast than a slow rate context and more initial responses were given the longer the critical sounds. The effect of rate was larger at longer durations of the continuum. Critically, for both trial types the interaction between rate and experiment was significant. The effect of rate was stronger in Experiment 2 than it was in Experiment 3, hence distal rate context modulated the use of immediate context. For s-trials the additional interaction between step and experiment suggests that listeners responded more categorically in Experiment 2 than in Experiment 3.

	s-trials		t-trials	
Factors	*b*	*p*	*b*	*p*
rate	-1.38	< .001	-1.15	< .001
step	0.07	< .001	0.11	< .001
experiment	-0.11	= .381	0.11	= .470
rate*step	-0.01	< .001	-0.01	< .05
rate*experiment	0.96	< .001	0.73	< .001
step*experiment	-0.02	< .001	--	--

Table 5. *Cross-experiment comparison of categorization data in Experiments 2 and 3. Factors were rate, step, experiment, and their interactions. Non-significant interactions that were eliminated from the model in only s-trials or t-trials are marked with "--".*

Discussion

Distal rate contexts influenced the use of speaking rate information that was directly adjacent to the target sounds and thus affected word recognition. The effect of proximal context in Experiment 3 was weaker than in Experiment 2. This is presumably because the proximal context in Experiment 3 was preceded by an opposite-rate distal context, whereas in Experiment 2 it was preceded by a distal context set to the same rate. This attenuation of the effect of immediate context was evident in the categorization task (both trial types) as well as in eye tracking (s-trials). Listeners did not perceive the proximal context as more extreme due to its contrast with the distal context but rather appeared to average the rate information over a longer stretch of time than the proximal context defined here.

Although distal context modulated the effect of the proximal context, listeners nevertheless still relied on speaking rate information from the proximal context to interpret the ambiguous sounds. In accordance with previous findings, both tasks showed that speaking rate context closer to the target had a greater influence on target perception than distal rate context. Note that in the eye-tracking analyses the effect of speaking rate did not reach significance but especially the t-trials showed a numerical effect similar to what was found in categorization.

Experiment 3 demonstrated an influence of distal context indirectly: it showed that distal rate context that had the opposite rate than the proximal context attenuated the effect of the proximal context. Experiments 4 and 5 further explored the influence of distal context on the word recognition process. For a direct assessment of effects of distal contexts on word segmentation, in Experiments 4 and 5 the proximal context was made uninformative for the disambiguation of the critical sounds. In this way any rate effects could be attributed to distal context. The questions were, first, whether listeners would use the distal context to disambiguate the critical sounds and, second, whether, in comparison to the effect of the proximal context in Experiment 2, the length of the distal context could compensate for its distal location.

Experiment 4

Experiment 4 asked whether distal rate context can directly modulate lexical competition of the target words. To maximize chances of detecting an effect of distal context, two manipulations were introduced in the preceding context. First, the context immediately adjacent to the targets was set to a normal speaking rate, that is, to the same rate as the target and the following context. Since the critical sounds were ambiguous at a normal speaking rate, the proximal context should be uninformative for word segmentation. Second, we used longer distal contexts than in Experiment 3 (see Figure 1). If listeners use these distal contexts to disambiguate the critical sounds they should be more likely to perceive the targets as initial following a fast than a slow rate context and they should be more likely to perceive the targets as non-initial following a slow than a fast rate.

A comparison to Experiment 2 where distal and proximal context had been rate-manipulated together should further provide insight into the role of context location relative to the target. Note that the preceding carrier sentence in Experiment 2 was shorter than in Experiment 4 but there the speaking rate context was directly adjacent to the critical sounds. Dependent on condition (i.e., s-trials, t-trials, initial - non-initial) the context in Experiment 2 was between 340 ms and 866 ms long whereas the distal context in Experiment 4 was between 935 ms and 1955 ms long. If the rate effect is stronger in Experiment 2 than in Experiment 4 this would suggest that the location of the rate context is crucial. Proximal context would be more important in word recognition than more distal context. But if the effect of rate in Experiment 4 is stronger than or equal to that in Experiment 2, this would suggest that there is a tradeoff between amount and location of rate context.

Methods

Participants

30 participants from the same population as in the previous experiments but who did not take part in any of the previous tasks received a small payment for their services.

Materials

The same speaker as in the previous experiments recorded a subset of the target words (x) from the s-trials and t-trials in longer context sentences. The new carrier sentences were 'Tijdens de lange vergadering heeft ze wel eens x gezegd' ("During the long meeting she once said x") for s-trials and 'Tijdens de lange vergadering heeft ze nooit x gezegd' ("During the long meeting she never said x") for t-trials. Note that due to Dutch syntactic structure constraints the sentences used in the previous experiments could not be lengthened by adding more material at their beginnings. One token of each new sentence was selected by matching the durations of the proximal context ('wel eens' and 'nooit') and the target words with the respective durations in the previously used sentences. This ensured comparable speaking rates across experiments. The distal contexts of the selected new sentences were again speeded up to 66% or slowed down to 133% of their original durations for fast and slow rate conditions. For the proximal context which was left at a normal rate, tokens of 'wel eens' and 'nooit' were taken from the normal rate condition in Experiment 1. These had durations of 195 ms ('wel eens') and 185 ms ('nooit') respectively. The critical sounds, the continua in the categorization task, the target words, and the following context were the same as in the previous experiments. All other aspects of the stimuli were identical to the previous experiments.

Procedure

Eye tracking

The timing of the previous experiments was slightly adapted. Due to the length differences of the new carrier sentences in the fast and slow rate conditions (i.e., 1169 ms vs. 2155 ms for s-trials and 1116 ms vs. 2066 ms for t-trials) an equal preview time from display onset to the critical sounds of 1800 ms was impractical. The different amounts of delay between display onset and start of the auditory stimuli could have led listeners to predict

the upcoming rate and thereby could have influenced its effect. Therefore the auditory stimuli were presented at a fixed interval of 300 ms after display onset. This left listeners even in the shortest sentence condition (i.e., fast t-trials) with sufficient time to read the four short words on the displays before the relevant acoustic information about the targets came available.

Categorization

To keep the categorization experiment with the long context sentences at the same length as the previous categorization experiments, each combination of speaking rate and step of the continuum was repeated 8 times instead of 10 times. The preview time of the response options relative to the start of the auditory stimulus was shortened from 600 ms to 200 ms. The long sentences and the invariant positions of response options on the screen provided participants with sufficient time to prepare their responses.

Results

Eye tracking

Data from eighteen trials (0.4%) were excluded from the analyses due to fixations outside the screen. On 93 trials (2.07%) participants clicked on words other than the target (67 trials, 1.49%) or outside the fields defining fixations to be on the words (26 trials, 0.58%). As before, all trials containing the targets 'tillen' and 'triest' were excluded from further analyses. Figure 7 plots fixation proportions to target, competitor, and the average of the two distractors over time for fast and slow rate from the onset of the critical sounds. Since the immediate context was kept at a normal speaking rate (i.e., the same rate as the target and following context), fast and slow here refer to the rate of the distal contexts.

Results showed an effect of distal rate context for initial and non-initial targets for the s-trials but not for the t-trials (see Table 2). For s-initial trials the target-competitor fixation difference was larger following a fast than a slow distal context. As expected, for not-s-initial trials the opposite was the case.

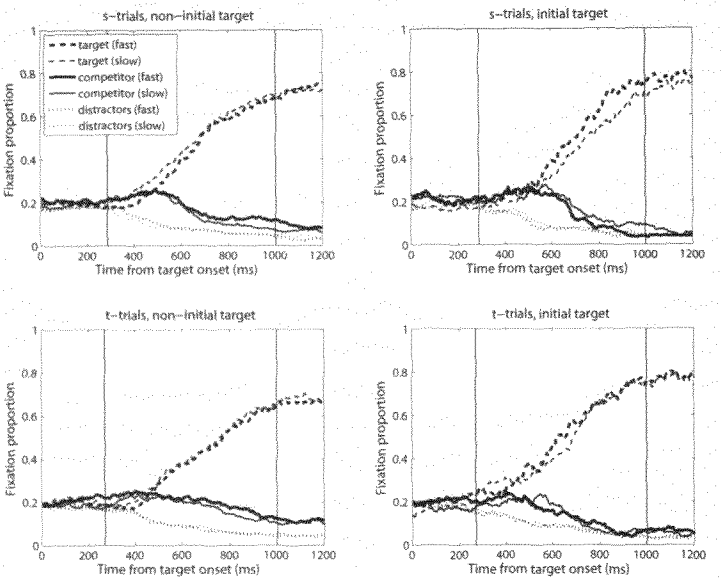

Figure 7. *Fixation proportions over time in the eye-tracking task in Experiment 4 to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines) following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. Fast and slow rate are coded for the rate of the long distal context. The immediate context had the same rate as the target words. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1000 ms after the onset of the critical sounds.*

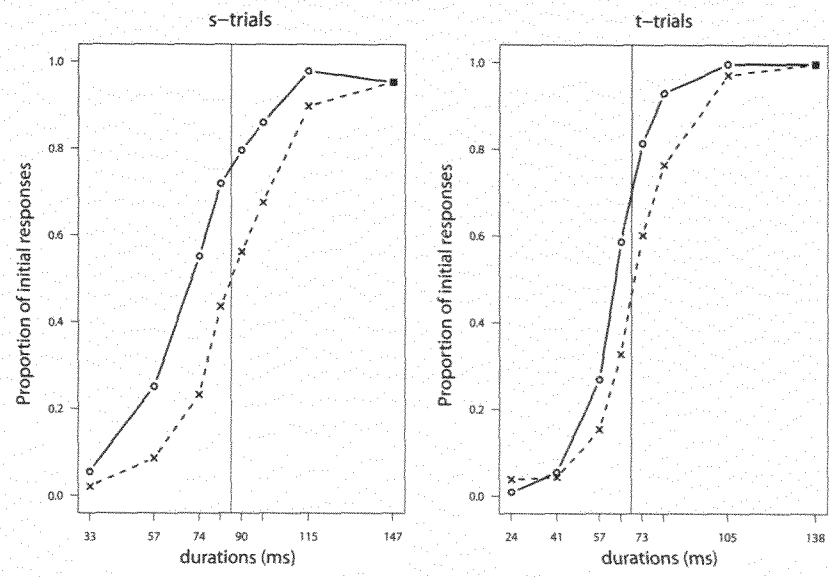

Figure 8. *Proportion of initial responses for s-trials and t-trials in the categorization task in Experiment 4. Duration of [s] and closure duration of [t] were varied on a subset of 8 steps from the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment. The long distal context was rate manipulated. The immediate context had the same rate as the target words.*

Categorization

Figure 8 shows categorization data for s-trials and t-trials along the duration continuum for fast and slow distal rate contexts. For s-trials main effects of rate and step were found (see Table 3). More s-initial responses were given following a fast than a slow distal rate context and more initial responses were given the longer the s-durations. For t-trials main effects of rate and step, as well as an interaction between these two factors were found. More t-initial responses were given following a fast than a slow distal context and more t-initial responses were given at longer step durations. The interaction suggests that the effect of rate was stronger for longer closure durations.

Cross-experiment comparisons

Eye tracking

In order to investigate the effect of the amount and location of rate contexts, Experiment 4 was compared to Experiment 2. As shown in Table 4, for the eye-tracking data the critical interaction between rate and experiment was significant for s-initial but not for t-initial targets. For s-initial targets, the effect of proximal rate context in Experiment 2 was stronger than the longer but distal rate context in Experiment 4. Further, main effects of rate were found for the s-initial, not-s-initial, and not-t-initial conditions. In the initial condition a fast rate context increased the fixation difference between target and competitor relative to a slow context. Slow contexts supported the recognition of non-initial targets.

Categorization

For the categorization data (see Table 6), the analysis of s-trials showed significant main effects of rate and step (more initial responses following a fast rate and the longer duration of [s]), as well as an interaction between these two factors (i.e., the effect of rate was stronger at longer s-durations). The critical two-way interaction between rate and experiment was not significant. Instead, the three-way interaction between rate, step, and experiment was significant. The difference in rate effect between Experiment 2 and Experiment 4 varied as a function of step. Figures 3 and 7 suggest that the rate effect was stronger in Experiment 2 than in Experiment 4 at the long steps of the continuum.

For t-trials main effects of rate and step were found, as well as an interaction between step and experiment (see Table 6). More initial responses were given following a fast than a slow context and the longer the closure duration. Listeners' responses were more categorical in Experiment 4. As for s-trials, the three-way interaction between rate, step, and experiment was significant. The stronger categorical perception in Experiment 2 than Experiment 4 was more evident for the fast than the slow rate condition. The overall effect of rate did not to differ between the categorization experiments.

	s-trials		t-trials	
Factors	*b*	*p*	*b*	*p*
rate	-1.38	< .001	-1.18	< .001
step	0.07	< .001	0.11	< .001
experiment	-0.08	= .574	-0.32	= .091
rate*step	-0.01	< .01	-0.01	= .098
rate*experiment	0.20	= .109	0.14	= .358
step*experiment	0.002	= .473	0.01	< .05
rate*step*experiment	0.01	< .05	-0.03	< .05

Table 6. *Cross-experiment comparison of categorization data in Experiments 2 and 4 evaluating the effects of rate, step, experiment, and their interactions.*

Discussion

Experiment 4 showed that distal rate context can directly modulate the perceived duration of the critical sounds. Whereas Experiment 3 showed an indirect effect of distal context on word recognition by modulating the effect of proximal context, in Experiment 4 the proximal context was not informative for the interpretation of the critical sounds. Effects of speaking rate thus reflect directly the use of distal context information. This finding was consistent for both types of tasks (although in eye tracking it was only found for s-trials). Compared to Experiment 3, it thus appears that informativeness of the immediate context is critical to the use of distal rate context in word recognition. Note that the use of distal context was not a result of task strategies. Additional analyses showed that trial number (i.e., trial rank order) in Experiment 4 could not predict the results nor interacted with the effect of rate, hence listeners did not learn during the experiment that informative cues were located further away from the target.

In addition to the change in informativeness of the proximal context, the distal rate context in Experiment 4 was longer than in the previous experiments. We asked whether the longer distal context in Experiment 4 would compensate for the fact that it was distal compared to the shorter but proximal context in Experiment 2 (see Figure 1). This comparison revealed interesting differences between tasks. Even though the proximal context in Experiment 4 was uninformative for the segmentation of the target word, its presence attenuated the effect of distal context on the interpretation of s-initial targets in the eye-tracking task. That is, in eye tracking proximity was more important than the amount of

context. In the categorization task no difference between Experiments 2 and 4 was found. That is, in categorization the amount of distal context can compensate for its distal location.

These results suggest that listeners give less weight to distal rate context in their estimation of speaking rate during the word recognition process than in categorization. The reason for this may be the difference in processing time available in the two tasks. During the word recognition process, as revealed in eye-tracking, listeners rely on a continuously updated estimate of rate in resolving the ongoing lexical competition process. In categorization, listeners have more processing time available before a response is initiated. Listeners therefore can re-visit the more distal context for rate information. This post-perceptual processing of distal context in categorization can also explain why a longer distal context can compensate for its distal location relative to a shorter but proximal context in categorization but not in eye tracking. Listeners have the opportunity to consider the rate of the additional longer context only in categorization.

The distal rate context in Experiment 4 appeared to have different effects on s-trials and t-trials in the eye-tracking task. Whereas effects of the long distal context were found for s-trials, no such effects could be found for t-trials. The reason for this may be that prosodic properties of the proximal context 'nooit' (i.e., in t-trials) were stronger than those of 'wel eens' (i.e., in s-trials). In spontaneous speech, 'wel eens' is more likely to be reduced than 'nooit'. So even though the only sentence accent was placed on the targets, a prosodically strong proximal context ('nooit') may attenuate effects of distal context more than a weak proximal context ('wel eens') would do. Note, however, that this hypothesis cannot be tested easily since, for example, a shift of sentence accent to the proximal context 'wel eens' would still leave 'eens' as a weak syllable forming the directly adjacent context. An alternative explanation for this difference between s- and t-trials may be that the cue of silence for the [t]-closure may be more difficult to interpret than the frication of [s] since the absence of signal may lead to uncertainty about what had been perceived. The fact that in Experiment 2 rate effects in the eye-tracking tasks were found for both cues, however, does not speak in favor of this explanation. In addition, the results for both cues were very similar in categorization.

In summary, Experiment 4 showed that listeners use distal context information in the perception of word boundaries. Whereas in categorization the effect of the long distal context was as strong as the directly adjacent context in Experiment 2, in eye tracking an effect of distal context was present for s-trials but attenuated by the proximal context. What cannot be decided from Experiment 4 is whether the amount of distal context contributed to

its effect. Experiment 5 therefore explored whether listeners can also use rate information from a short distal context.

Experiment 5

Experiment 5 investigated the role of the amount of distal rate context on the perception of the ambiguous sound sequences. In Experiment 4 long carrier sentences were used in order to maximize the chances of detecting an effect of distal context. Experiment 5 examined whether a short distal context (i.e., dependent on condition between 217 ms and 617 ms) is sufficient for listeners to disambiguate the critical sounds. In Experiment 5 we therefore combined the short distal contexts from Experiment 3 with the neutral-rate proximal contexts from Experiment 4 (see Figure 1). Cross-experiment comparisons then addressed whether a long distal rate context (Experiment 4) would yield stronger rate effects than a short distal context (Experiment 5). Further, a comparison between Experiment 5 and Experiment 3 followed up on the question about the role of informativeness of intervening proximal context on effects of distal contexts. Both Experiment 3 and 5 presented the same preceding sentences, and the rate of the distal contexts was manipulated in the same fashion. What distinguished the two experiments was the manipulation of the proximal contexts. Whereas in Experiment 3 the proximal contexts could be used for the disambiguation of the critical sounds, proximal contexts were not informative in Experiment 5.

Methods

Participants

30 new participants from the same population as in the previous experiments received a small payment for their services.

Materials, Procedure, Design

The distal context in Experiment 4 was replaced by the distal context used in Experiment 3. In all other aspects the experiment was identical to Experiment 4. That is, the distal context was rate-manipulated while the proximal context was always presented at a normal rate. The proximal context was thus not informative for the interpretation of the critical sounds.

Results

Eye tracking

98 trials (2.18%) were excluded from further analyses due to fixations outside the screen (10 trials, 0.22%) or due to participants' clicks outside the target area (67 trials, 1.49%, clicks on competitor; 21 trials, 0.47%, clicks outside the regions defining the words). As before, trials with the targets 'tillen' and 'triest' were also excluded. Figure 9 shows fixation proportions to targets, competitors, and the average of the two distractors over time for the two rate contexts from the onset of the critical sounds. It seems that s-initial targets are recognized better following a fast than a slow distal rate context. As evident from Table 2, however, this effect was not significant. A short distal rate context did not modulate lexical competition in any of the target conditions.

Categorization

Figure 10 shows categorization data for s-trials and t-trials along the duration continuum for fast and slow distal rate contexts. For s-trials main effects of rate and step were found (see Table 3). More s-initial responses were given following a fast than a slow distal rate context and more s-initial responses were given the longer the s-duration. For t-trials a main effect of step, as well as an interaction between rate and step were found. More t-initial responses were given the longer the t-closure. The effect of rate differed as a function of step. The lack of a main effect of rate suggests that no rate effect was present for the duration that had been used in the eye-tracking experiment.

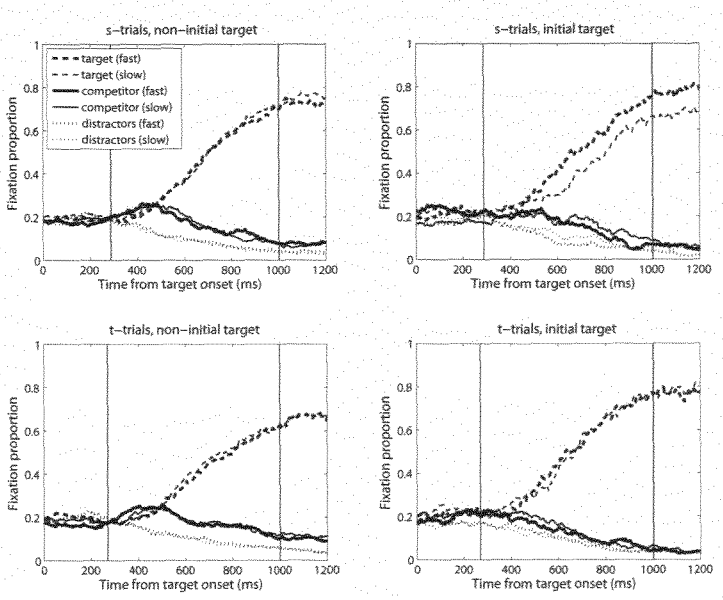

Figure 9. *Fixation proportions over time in the eye-tracking task in Experiment 5 to target (dashed lines), competitor (solid lines), and the average of the two distractors (dotted lines) following a fast (thick lines) or a slow (thin lines) rate context from the onset of the critical sounds. Fast and slow rate are coded for the rate of the short distal context. The immediate context had the same rate as the target words. The upper two graphs show initial and non-initial target conditions for the s-trials. The lower two graphs show initial and non-initial target conditions for the t-trials. Vertical solid lines mark the time window used for analyses, that is, from the offset of the critical sounds shifted by 200 ms until 1000 ms after the onset of the critical sounds.*

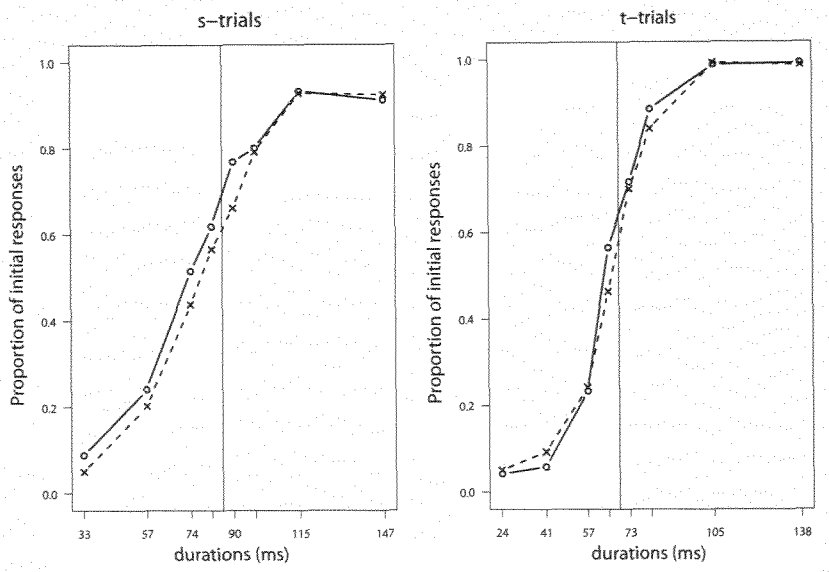

Figure 10. *Proportion of initial responses for s-trials and t-trials in the categorization task in Experiment 5. Duration of [s] and closure duration of [t] were varied on a subset of 8 steps from the duration continuum in Experiment 1. The vertical line marks the duration that was used in the eye-tracking experiment. The short distal context was rate manipulated. The immediate context had the same rate as the target words.*

Cross-experiment comparisons

Experiment 5 vs. Experiment 4: Amount of distal context

Experiment 5 was compared to Experiment 4 to assess whether a shorter distal context would lead to smaller rate effects than a longer distal context. In both experiments the proximal context was uninformative about rate. Table 4 summarizes the eye-tracking results. For none of the conditions was an interaction between rate and experiment found. Only for s-initial trials was the main effect of rate significant. Distal rate context thus affects online word recognition in at least one condition but the amount of distal context appears not to matter.

The categorization data (see Table 7) showed for s-trials main effects of rate and step, as well as an interaction between step and experiment. More s-initial responses were given following a fast than a slow rate context and more s-initial responses were given the longer

the s-duration. Responses were more categorical in Experiment 4 than in Experiment 5. Moreover, rate interacted with experiment suggesting that the effect of rate was stronger in Experiment 4, that is, the amount of distal context rate influenced the effect of rate on perception. For t-trials main effects of rate, step, and an interaction between these two factors were found. More t-initial responses were given following a fast than a slow rate context, and the longer the [t]-closure. The effect of rate was stronger the longer the steps of the continuum. Critically, the interaction between rate and experiment was significant. The effect of rate was larger in Experiment 4 than in Experiment 5, hence in categorization a longer distal rate context led to stronger influences on listeners' behavior.

		s-trials		t-trials	
	Factors	*b*	*p*	*b*	*p*
Exp.4 vs. Exp.5	rate	-1.19	< .001	-0.99	< .001
	step	0.07	< .001	0.12	< .001
	experiment	0.11	= .457	0.27	= .267
	rate*step	--	--	-0.03	< .001
	rate*experiment	0.93	< .001	0.77	< .001
	step*experiment	-0.02	< .001	--	--
Exp.3 vs. Exp.5	rate	0.42	< .001	0.41	< .001
	step	0.05	< .001	0.10	< .001
	experiment	0.11	= .392	-0.25	= .346
	rate*step	0.01	< .001	0.01	= .187
	rate*experiment	-0.62	< .001	-0.61	< .001
	step*experiment	--	--	0.01	= .062
	rate*step*experiment	--	--	-0.02	< .05

Table 7. *Cross-experiment comparisons of categorization data in Experiment 5 with Experiment 4 and Experiment 3, evaluating the effects of rate, step, experiment, and their interactions. Non-significant interactions that were eliminated from the model in only s-trials or t-trials are marked with "--".*

Experiment 5 vs. Experiment 3: Informativeness of proximal context

The comparison of Experiment 5 to Experiment 3 investigated whether the informativeness of the proximal rate context affected the use of short distal rate contexts in word recognition. In Experiment 5 the proximal context had a neutral speaking rate, that is, the same rate as the target word and following context. Since the critical sounds were ambiguous at a neutral rate the proximal context was not informative about target interpretation. In Experiment 3 the proximal context had the opposite rate than the distal context and thus provided disambiguating information for the interpretation of the critical sounds. Note that in contrast to the analyses of Experiment 3 where the terms fast and slow rate were used with regard to the proximal context, for the present comparison the data were recoded to reflect slow vs. fast distal contexts.

Results of the eye tracking analysis showed that the informativeness of the proximal context did not modulate effects of distal context (see Table 4). Short distal contexts, however, were not sufficient for listeners to disambiguate the critical sounds. None of the factors nor the critical interaction between rate and experiment was significant in any of the conditions.

The comparison of the categorization data (see Table 7) showed for s-trials main effects of rate and step, and an interaction between these two factors. The target was perceived as s-initial more often following a slow than a fast distal rate context and the longer the [s]-duration. This context effect might appear to be in the wrong direction, but this is not the case because it reflects the use of **proximal** context in Experiment 3. The slow distal contexts in that experiment were paired with fast proximal contexts, so, in line with all other rate effects with [s] reported here, these fast proximal contexts were followed by more s-initial interpretations. In addition, the effect of rate was stronger the longer the steps of the continuum. Critically, the interaction between rate and experiment was also significant. The effect of rate was stronger in Experiment 3 than in Experiment 5. This result again should be interpreted as reflecting the use of proximal context in Experiment 3.

For t-trials main effects of rate and step were found. Participants gave more t-initial responses following a slow than a fast distal rate context. As for s-trials this appears to be because of the use of proximal context in Experiment 3. The critical interaction between rate and experiment was also significant. The effect of rate was stronger in Experiment 3 than in Experiment 5. In addition a three-way-interaction between rate, step, and experiment was

found. The effect of rate differed as a function of step in Experiment 5 but not in Experiment 3.

Discussion

Experiment 5 demonstrated that in the categorization task a short distal context can be used to recognize the targets of s-trials and t-trials. Listeners indeed appear to derive rate information from a time window that is somewhat larger than the 185 ms and 195 ms proximal contexts that we used. The assumption of a 300 ms time window over which according to some accounts (Newman & Sawusch, 1996; Sawusch & Newman, 2000) rate information is averaged thus seems to be supported. In comparison to the effect of the long distal context in Experiment 4, however, the short distal context had a much smaller effect. This suggests that either the time window over which rate information is averaged is larger than 300 ms (Newman & Sawusch, 1996; Sawusch & Newman, 2000), or that there is some other mechanism that allows a longer distal context to have a larger effect. Note that Wayland et al. (1994) suggested different effects of long vs. short preceding rate contexts.

The present series of experiments reveals that effects of the amount of rate context appear to be restricted to offline tasks. No advantage of a longer distal context was found in eye tracking, neither in the comparison between the proximal context in Experiment 2 and the long distal context in Experiment 4, nor in the comparison between the long distal context in Experiment 4 and the short distal context in Experiment 5. These results support the hypothesis that more rate context information can be taken into account when more processing time is available (i.e., for post-perceptual decisions in categorization compared to online processing in eye tracking).

Concerning the influence of the informativeness of the proximal context on the use of distal rate information, the categorization results clearly confirm the importance of the proximal context in word recognition. Although it seems that the effect of distal context was stronger in Experiment 3 (informative proximal context) than in Experiment 5 (not informative proximal context), the direction of the rate effect provides the crucial result. Specifically, in the cross-experiment comparison the main effect of rate indicated that more initial responses were given following a slow than a fast distal context. But, as argued above, this reflects the greater use of (opposite-rate) proximal context in Experiment 3. When interpreted this way, the result is consistent with all other effects observed here (i.e., more initial responses after fast proximal contexts). This suggests that if the proximal context is

informative for the interpretation of the critical sounds, listeners use the proximal context rather than the distal context.

General Discussion

A series of categorization and eye-tracking experiments investigated how speaking rate information from a preceding sentence context influences the perception of durational cues to word boundaries. Ambiguous Dutch word sequences of the type 'wel eens (s)peer' (s-trials) and 'nooit (t)rap' (t-trials) were presented in rate-manipulated context sentences such that the boundary sounds [s] and [t] could either be interpreted as pre-juncture phonemes or, in addition, as the initial sounds of the target words. The juncture phonemes [s] and [t] were chosen to test whether speaking-rate effects depend on the type of durational cue in the signal, that is, frication noise in [s] vs. silence during the closure of [t]. Previous studies on word segmentation showed that the longer the boundary sounds are, the more likely these sounds are perceived as target initial (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, et al., 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli et al., 2003; Tabossi et al., 1995). Experiment 1 therefore first asked with a categorization task whether the duration of boundary sounds and thus their position were perceived relative to the rate of the preceding context. As predicted, the critical sounds were perceived as target initial more often following a fast than a slow context sentence. Results for the two boundary sounds [s] and [t] were largely similar despite the differences in the acoustic characteristics of their durational cues.

Then two major questions were addressed: First, when during word recognition do listeners use speaking rate information? Second, where relative to the critical sounds must this rate information be in order to be used? Critically, two different tasks provided insight into how speaking rate information modulates lexical competition during the word recognition process (eye tracking) vs. how speaking rate affects listeners' responses when more processing time is available (categorization).

Experiment 2 introduced an eye-tracking task to examine rate effects during the word recognition process. Listeners interpret durational information in relation to preceding rate context as the signal unfolds. Speaking rate thus appears to be taken into account prelexically to incrementally inform the evaluation of upcoming durational cues. This is in line with previous categorization studies on the use of rate information which already suggested that rate information is taken into account during an early phase of speech

processing. Listeners, for example, tended to use rate information from sources other than the target speaker (Sawusch & Newman, 2000; Newman & Sawusch, 2009) and even from locations other than the target speakers' location (Newman & Sawusch, 2009). That is, speaking rate information is taken into account before other low-level processes such as stream segregation occur. Moreover, speaking rate information cannot be ignored (Miller & Dexter, 1988). When listeners responded quickly, before they heard the complete stimulus, they consistently treated the vowel following the target phoneme as short. Listeners thus used all information available at the time of responding. This is in line with other evidence showing that word recognition is optimal and incremental (Dahan, Magnuson, Tanenhaus, & Hogan, 2000; Norris & McQueen, 2008; Reinisch et al., 2010, in press; Tanenhaus et al., 1995). The present results add direct evidence that, as the speech signal is processed over time, the rate estimate is continuously updated and applied during low-level speech processing.

Previous research on the effects of speaking rate suggested that rate information that is close to the target has a larger influence on target perception than distal speaking rate context (e.g., Summerfield, 1981). The time window within which the rate of a context is considered has been estimated to be about 300 ms (Newman & Sawusch, 1996; Sawusch & Newman, 2000). As the acoustic signal unfolds, the estimation of speaking rate has to be updated moment by moment and in this process more weight is given to newly-arriving information. Even though most previous speaking rate studies focused on rate information in the close vicinity of the target sound, distal rate information has been shown to influence speech perception as well (e.g., Kidd, 1989; Summerfield, 1981; Wayland et al., 1994). Experiments 3-5 further investigated under which circumstances more distal rate information modulates the word recognition process. Since the range of the proximity effect has been estimated between 250 and 300 ms (Newman & Sawusch, 1996; Sawusch & Newman, 2000) we defined proximal context as maximally 250 ms (the slow proximal context in Experiment 3). In this way a part of our distal contexts was sufficiently close to the critical sounds to potentially have an effect. The question was whether additional manipulations of distal contexts would affect the use of proximal context and whether distal contexts alone were sufficient to influence target perception.

When the rate of the distal context had the opposite rate than the proximal context, distal context attenuated the effect of proximal context (Experiment 3). Distal context thus can affect word recognition. Distal rate context can also have a direct effect on word recognition. In Experiments 4 and 5 the proximal context was set to a neutral rate (i.e., the intervening proximal context was of equal duration for fast and slow distal contexts, see

Figure 1) and thus was not informative for the interpretation of the target words. Listeners nevertheless interpreted the targets relative to the earlier distal context (Experiments 4 and 5). Given our choice of duration for proximal contexts these effects support the 300 ms time-window hypothesis.

Additional results, however, suggest that the 300 ms time-window account may be too simplistic. In the categorization tasks, the amount of a distal context appeared to compensate for its distal location (Experiment 4 vs. Experiment 2) and a longer distal context led to stronger rate effects than a short distal context (Experiment 4 vs. Experiment 5; see Figure 1). The amount of context information thus appears to matter. Similar effects of amount of distal context have been found in a categorization study on the role of distal rhythmic context (i.e., pitch and duration) in the segmentation of compound words (Dilley & McAuley, 2008). In this study sequences of four words had to be categorized as either consisting of two compound words (e.g., 'footnote' 'bookworm'), or of a compound word surrounded by two monosyllabic words ('foot' 'notebook' 'worm'). Depending on the duration and pitch patterns of additional preceding context words listeners grouped the target sequences according to the preceding rhythmic pattern even though acoustically the last three target words were kept constant across conditions. Importantly, the effect of context was stronger if it was long (i.e., the full context) than if it was truncated. Listeners thus not only use various types of distal phonetic context information (i.e., rhythmic information in Dilley and McAuley study, speaking rate in the present series of experiments) but the amount of this distal context modulates the strength of the effect. In categorization tasks listeners thus appear to be able to use all previous information, resulting in a cumulative effect of distal context.

Since Experiment 2 demonstrated that listeners immediately interpret upcoming durational information relative to the speaking rate of the preceding context, an important question was how much rate information listeners keep available during online speech perception. Given the need for a continuous update of speaking rate estimate during the word recognition process, the effect of rate context in eye tracking may not be cumulative as has been suggested for the categorization tasks. Indeed in online processing rate information that is closer to the target is much more important than more distal context. In contrast to the categorization results, the distal contexts in eye tracking showed weaker effects, and also the amount of distal context appeared to play a lesser role. The shorter but proximal context in Experiment 2 had stronger effects than the long but distal context in Experiment 4. Moreover, the long distal context in Experiment 4 did not show an advantage over the short distal context in Experiment 5. Proximity is thus a crucial factor in the online use of rate

information. Nevertheless, listeners did use information from the distal rate contexts during word recognition. The effects of proximal context in Experiment 3 were attenuated by the distal context, and, importantly, the long distal context in Experiment 4 affected the interpretation of the boundary sound [s]. It is thus not the case that rate information in distal contexts has no effect on online word recognition.

In summary, speaking rate information is used during the word recognition process to prelexically evaluate durational cues. There were no substantial differences in rate effects due to differences in the acoustic characteristics of these cues. Although context information that is proximal to the target had stronger effects than distal context, listeners were able to use distal context information to recognize the target words. Effects of distal context were especially prominent when the distal context was long and the proximal context was uninformative for target interpretation. In addition, listeners used distal rate information more when time was available for them to give an explicit response about what they perceived than during online word recognition. This suggests that, while offline tasks such as categorization can reveal the limits of the perceptual system, they do not necessarily show how speech information is used in the early stages of word recognition. With respect to online processing, it appears that local rate information carries the most weight. Listeners thus efficiently evaluate durational information to word boundaries by immediately taking into account speaking rate information estimated primarily from the proximal preceding context.

Appendix

List of targets and competitors for s-trials and t-trials

initial targets							
s-trials				t-trials			
Dutch		English translation		Dutch		English translation	
target	competitor	target	competitor	target	competitor	target	competitor
slaaf	laars	slave	boot	tent	eng	tent	scary
slank	lap	slender	piece	test	engel	test	angel
slepen	lenen	drag	lend	thee	edel	tea	noble
smerig	meten	dirty	measure	thema	eeuwig	theme	eternal
speler	peper	player	pepper	tocht	offer	journey	victim
spoelen	poes	wash	cat	toetsen	oester	test	oyster
sleutel	leus	key	slogan	toeval	oever	coincidence	shore
spotten	pols	spot	wrist	ton	omgang	cask	contact
schade	chaos	damage	chaos	tong	order	tongue	order
schelen	chemisch	differ	chemical	traag	raar	slow	odd
sfeer	feest	atmosphere	party	traan	raad	tear	advice
slikken	lichten	swallow	light up	tragisch	raadsel	tragically	riddle
sneeuw	neef	snow	nephew	trappen	ramp	kick	disaster
slim	lift	smart	elevator	tres	rector	braiding	rector
smaken	maag	taste	stomach	trend	rechts	trend	right
stress	trechter	stress	funnel	triest	ring	sad	ring
steunen	teugel	support	rein	truc	rum	trick	rum
stekker	technisch	plug	technical	trui	ruilen	shirt	change

non-initial targets							
s-trials				t-trials			
Dutch		English translation		Dutch		English translation	
target	competitor	target	competitor	target	competitor	target	competitor
pad	spalk	path	splint	aak	taai	barge	tough
pan	spar	pan	spruce	aal	tanig	eel	tawny
peer	speen	pear	pacifier	aard	tafel	earth	table

peil	spijt	level	regret	actie	tachtig	action	eighty
pellen	sperren	peel	blocks	al	tak	already	branch
pier	spies	worm	skewer	as	tact	ash	tact
pil	spits	pill	peak	eelt	teek	horny skin	tick
pit	spin	seed	spider	eer	tevens	honor	besides
poel	spoed	pool	hurry	egel	tegen	hedgehog	against
pook	spoor	gearstick	track	el	term	yard	term
port	spons	harbor	sponge	eren	telen	honor	breed
pot	spon	jar	tap	iepen	typisch	elms	typical
pul	spurt	tankard	spurt	ijl	tijm	haste	thyme
taak	staaf	task	bar	innen	tippen	collect	tip
taal	staan	language	stand	issue	ticket	issue	ticket
taart	staand	cake	standing	oog	toon	eye	tone
tal	stad	number	town	oost	toorn	east	rage
tam	staf	tame	staff	rand	trance	edge	trance
tand	stamp	tooth	kick	rap	tractor	quick	tractor
tap	stam	tap	trunk	rede	trema	reason	diaeresis
teen	steeg	toe	lane	rein	treil	clean	trawl
teil	stijf	washtub	stiff	rek	tref	elasticity	hit
teken	stelen	sign	steal	rekken	treffen	stretch	encounter
tel	ster	count	star	rem	tred	brake	step
tellen	stemmen	count	vote	reuzel	treuren	lard	mourn
tempel	stengel	temple	stalk	rib	tril	rib	vibration
tikken	stichten	tap	found	rillen	trimmen	shiver	exercise
tillen	stiekem	lift	secretly	roebel	troetel	rouble	darling
tip	stift	tip	felt-pen	roep	troef	call	trumps
toer	stoet	tour	crowd	rol	trom	role	drum
tof	stok	great	stick	rommel	troffel	junk	trowel
tol	stop	toll	stop	rooster	tropen	schedule	tropics
trant	stram	style	stiff	ros	trots	horse	pride
trekken	stremmen	pull	obstruct	rouwen	trauma	grieve	trauma
trip	strikt	trip	strict	ruk	truffel	jerk	truffle
tulp	stunt	tulip	stunt	urnen	turbo	urns	turbo

Summary and conclusions

Chapter 6

The aim of this thesis was to gain insights into the process of how listeners use temporal information to recognize spoken words in the continuous speech stream of their native language. As discussed throughout this thesis, the speech signal unfolds at a variable pace. Utterances can be spoken more quickly or more slowly. In addition, sound durations depend not only on their articulatory properties and speaking rate, but also on structural influences such as their position in a word's prosodic structure and the location of that word in an utterance. This thesis examined how listeners deal with these durational variations, and whether and how they can use this temporal variation to inform the word recognition process.

When listening to speech, listeners immediately use available segmental information to continuously update their hypotheses about the words they hear. Here I asked whether listeners also use suprasegmental information to recognize words. In particular, I investigated whether suprasegmental durational cues to lexical stress and to word boundary location are perceived relative to the speaking rate of the preceding context. The focus was on the processing of these cues over time, that is, online as the speech signal unfolds. Effects on online processing as measured with eye tracking were then compared to effects that also include post-perceptual processing as measured in tasks (i.e., categorization) that assess the result of the word recognition process. This comparison attempted to capture word recognition as it happens as well as assess the limits of the word recognition process. Further insights about the calculation of rate information in online vs. offline processing were sought by investigating speaking rate effects from proximal and distal contexts.

Visual-world eye tracking is an excellent method to monitor word recognition over time. Previous studies using this paradigm showed the incremental uptake of segmental information during word recognition and demonstrated listeners' sensitivity to fine phonetic detail in the segmentation of words (Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). The eye-tracking experiments reported here tested listeners' uptake of

suprasegmental cues to lexical stress and the use of speaking rate context over time. All experiments used printed words for the visual display (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Using printed words instead of pictures made it possible to use a larger number of target-competitor pairs than could be used if depictable words had to be found. In addition, printed words selectively tap into phonological processing (Huettig & McQueen, 2007).

Chapter 2: Early use of lexical stress information

The experiment described in Chapter 2 investigated when during the word recognition process Dutch listeners use cues to words' lexical stress patterns. Lexical stress provides temporal information on a larger time scale than segment durations since lexical stress patterns are distributed over the syllables of a word. The processing of stress information in a word thus involves the recognition of alternating degrees of stress. Unlike in English, where unstressed vowels get reduced, Dutch lexical stress is mainly marked suprasegmentally, that is, by duration, pitch, amplitude, and spectral tilt. Conducting the experiment in Dutch thus made it possible to investigate effects of suprasegmental stress information without interference from segmental stress effects. Previous studies suggested that Dutch lexical stress can be beneficial for word recognition (Cutler & Pasveer, 2006; van Heuven & Hagman, 1988). Using stress information significantly reduces the number of embedded words (Cutler & Pasveer, 2006). It also allows for an earlier disambiguation of words (van Heuven & Hagman, 1988). Dutch listeners indeed use suprasegmental stress information in word recognition (e.g., Cooper, Cutler, & Wales, 2002; Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005; van Heuven, 1988; van Leyden & van Heuven, 1996). For example, in a fragment priming experiment (van Donselaar et al., 2005), matching stress patterns between primes and targets facilitated target recognition even if the primes were only one syllable in length (e.g., the prime 'OC-' facilitates the recognition of 'OCtopus', "octopus"; capitals indicate stress). Mismatching primes, however, had to be of at least two-syllable length in order to inhibit target recognition (e.g., 'okTO-' 'OCtopus'). Mismatching one-syllable primes neither facilitated nor inhibited target recognition. Previous studies thus suggested that the amount of stress information listeners have available (i.e., one or two syllables) influences word recognition. None of the previous studies, however, was able to demonstrate when exactly during the word recognition process listeners take stress information into account.

The experiment reported in Chapter 2 therefore employed an eye-tracking experiment to monitor stress-modulated lexical competition as the speech signal unfolds. Dutch target-competitor pairs were selected such that they overlapped segmentally for at least their initial two syllables but differed in lexical stress placement. The word pairs contrasted stress placement on either their first vs. second syllable (e.g., 'OCtopus' - 'okTOber'; "octopus" - "October", 1-2 contrast) or on their first vs. third syllable (e.g., 'DIameter' – 'diaMANT; "diameter" – "diamond", 1-3 contrast). Dutch words with primary stress on the third syllable have secondary stress on their initial syllable. Each word belonged to one of four conditions resulting from the combination of stress contrast (1-2 vs. 1-3 contrast) and stress location (initial vs. non-initial).

Results showed that listeners optimally use suprasegmental stress information as soon as it comes available. Listeners recognized the words based on their lexical stress patterns before segmental cues disambiguated the words. These results were similar for three out of the four conditions. Only words with primary stress on the third syllable were difficult to distinguish from their competitors on the basis of lexical stress cues alone. Since these words, however, have secondary stress on their initial syllable, and both words of the 1-3 contrast pairs have an unstressed second syllable, their initial two syllables were perceptually highly similar. Therefore words with primary stress on the third syllable were expected to be hardest to recognize on the basis of only stress information. Furthermore, initially stressed words ('OCtopus') competed more strongly for recognition than non-initially stressed words ('okTOber') – independently of which word was the target. These results demonstrated that stress information is used during the word recognition process and that stress contrast (1-2 vs. 1-3 contrast) and stress location (initial vs. non-initial stress) modulate the words' competition for recognition.

The results were relevant for another discussion in the stress-perception literature. Previous studies on stress perception reported a bias for listeners to perceive word-initial stress (van Leyden & van Heuven, 1996). This bias was mostly attributed to lexical statistics showing that more words in Dutch (and English) have word-initial than non-initial stress (van Heuven & Hagman, 1988). In the experiment reported here, early differences in the competition patterns between initially and non-initially stressed targets were found. Initially stressed words competed more strongly for recognition than non-initially stressed words. This suggests that the previously reported initial-stress bias is not purely driven by statistics. Rather, the stronger competition of initially stressed words is signal driven. The presence of stress is more easily interpreted than the absence of stress. If a cue is perceived as absent, listeners may be unsure whether the cue was indeed absent or whether it had just been

missed. So although previously reported statistical effects may bias word recognition during a later phase of processing, there appears to be an early, signal-driven effect that leads listeners immediately to weigh the presence of initial stress cues more heavily.

A correlation analysis further investigated which acoustic cues to lexical stress listeners used during word recognition. The only cue that predicted listeners' fixations on the target words was duration. A larger difference between the durations of first vowels of initially and non-initially stressed words in a pair led to a larger difference in fixations on initially vs. non-initially stressed target words. This is in line with previous findings (e.g., Nooteboom, 1972; Sluijter and van Heuven, 1995, 1996) that suggested that duration is the most reliable lexical-stress cue in Dutch. Given this special status of duration as cue to stress, the question arose whether a syllable's degree of stress is interpreted relative to temporal context information, that is, relative to the speaking rate of a preceding utterance. This issue was addressed in Chapters 3 and 4. Chapter 2 showed that the temporal information about stress contained in a word itself is used as soon as it could be to constrain the word recognition process.

Chapter 3: Stress perception relative to speaking rate

The interpretation of temporal information crucially depends on the speaking rate of the utterance. Previous literature showed that **segmental** durational cues are interpreted relative to speaking rate (e.g., Allen & Miller, 2001; Miller, 1981; 1987; Miller & Dexter, 1988; Miller & Liberman, 1979; Miller & Wayland, 1993; Wayland, Miller, & Volaitis, 1994). When, for example, the word 'bin' is heard in a fast rate context, listeners are likely to report hearing 'pin'. Relative to the shortened segments in the fast rate context, the voice onset time (i.e., the time span between the release of the stop closure and the start of vocal-fold vibration) of the /b/ sounds long and thus leads to the perception of /p/. The most commonly used tasks in this research were categorization and goodness judgment tasks, both of which tap into perceptual and post-perceptual processing. Nevertheless, a number of studies suggested that rate effects occur during an early phase of processing (Miller & Dexter, 1988; Newman & Sawusch, 2009; Sawusch & Newman, 2000). For example, listeners use rate information from speakers other than the target speaker (Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000) even if this speaker is at a different spatial location (Newman & Sawusch, 2009). Speaking rate thus appears to be taken into account before other early processes such as perceptual grouping or stream segregation occur. Since duration is found to consistently cue lexical stress, and

speaking rate is likely to be used during an early phase of processing, speaking rate was expected to also affect listeners' immediate use of lexical stress information (as shown in Chapter 2). Lexical competition between initially and non-initially stressed words should be modulated by the speaking rate of the preceding context. Following a slow context the initial syllables of a word should sound relatively short (i.e., unstressed) and thus lead to interpretations of non-initial stress more often than following a fast context. This was tested in two eye-tracking experiments in Chapter 3.

These experiments employed the same task and materials as were used in Chapter 2. In the first experiment only the preceding context sentence was manipulated. The target words contained all cues to lexical stress (i.e., duration, pitch, amplitude). In this first experiment, however, the expected rate effects could not be found. Therefore the second experiment tested whether rate effects may have been disguised by the presence of stress cues on the targets other than durational cues. In the second experiment these other stress cues were neutralized. Pitch and amplitude cues on the initial two syllables of each word pair were set to their averaged values. The second experiment showed that when duration is the only cue to lexical stress, lexical competition is stronger than when more stress cues are available (as in the first experiment). Speaking rate, however, again did not show the expected shift in competition patterns. Rather, a fast speaking rate context led to faster target recognition independent of stress location. Nevertheless, duration was a sufficient cue for listeners to recognize words before segmental information came available. The experiments in Chapter 3 thus replicated the finding from the experiment in Chapter 2 that listeners immediately use of lexical stress information to recognize words.

The fact that duration was a sufficient cue for listeners to resolve lexical competition, however, could be a reason why speaking rate effects on lexical competition could not be found. Speaking rate may have been not strong enough to contribute additional information to the durational stress cues on the target words themselves. The results of Chapter 3 thus suggested that speaking rate does not affect the online processing of lexical stress information, as measured in eye tracking.

Chapter 4: Speaking rate affects uptake of lexical stress cues

To further investigate the relation between speaking rate effects and stress perception, the series of experiments described in Chapter 4 tested with a different task whether rate effects may depend on the presence of stress cues on the initial and non-initial syllables of target words. In a fragment categorization task, listeners had to decide which

word a two-syllable word fragment was taken from. The fragment was presented at the end of a fast or slow carrier sentence. A subset of words from the previous experiments was selected for manipulation. One word pair was taken from the 1-2 stress contrast ('Alibi' – 'aLInea'; "alibi" – "paragraph") and one word pair was taken from the 1-3 stress contrast ('CAvia' – 'kaviAAR', "guinea pig" - "caviar"). The duration of the initial or second vowel of the fragments was manipulated along a duration continuum while the respective other vowel was set to a duration that was perceptually ambiguous with respect to stress. In that way additional questions about rate effects on initial and non-initial syllables could be addressed.

Here robust effects of speaking rate on the perception of lexical stress patterns were found. Speaking rate affected the perception of stress on initial and second syllables of words from both stress contrasts (1-2 and 1-3 contrast). Rate effects did not depend on the perception of stress cues on the respective syllables. Second-syllable durations, although affected by rate, were not taken into account even when manipulated along the continuum. This supports the findings from Chapters 2 and 3 that stress information is used incrementally. Moreover, the influence of rate-modulated stress cues was larger on initial than on non-initial syllables. Together, the stronger rate effects on initial syllables and the small effect of second syllable duration suggest an incremental use of speaking rate context in the interpretation of durational cues to lexical stress.

Durational cues to lexical stress appeared to be perceived relative to the speaking rate of a preceding context when listeners' behavior was measured with a categorization task (Chapter 4) but not when the task tapped into the word recognition process (i.e., the eye-tracking task in Chapter 3). During the word recognition process listeners continuously update the acoustic information used for lexical access. Moreover, for stress perception, the relation of stress cues within a word may be more important than initial-syllable stress relative to the preceding context. In that way the presence of stress cues on the target words in eye tracking could disguise the effects of rate. In contrast, when listeners have time to process and maybe even reprocess the stimuli (and thus also the preceding context) as is the case in a categorization task, speaking rate is taken into account for the interpretation of durational cues. Even durational cues that are themselves informative about stress are interpreted relative to speaking rate. Nevertheless, stronger rate effects on initial than non-initial syllables in categorization suggested that speaking rate context is taken into account incrementally, as would be expected if speaking rate was evaluated during the early phases of processing. The experiments in Chapter 5 therefore further investigated the time-course of processing speaking rate information. When during word recognition is speaking rate

information taken into account? Chapter 5 also again addressed what role task differences could play with regard to the use of speaking rate information.

Chapter 5: Speaking rate effects on word segmentation

The series of experiments reported in Chapter 5 extended the topics of previous chapters in various ways. First, it investigated whether speaking rate effects generalize to another suprasegmental durational cue, namely duration as a cue to word boundary location. Ambiguous word sequences of the type 'wel eens (s)peer' and 'nooit (t)rap' ("once spear/pear", "never staircase/quick") were used to investigate listeners' perception of the last words (i.e., the targets) as [s] or [t]-initial or not. The longer the boundary sounds are, the more likely listeners interpret them as word-initial (e.g., Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, Liberman, Eccardt, & Pesetsky, 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Burani, & Scott, 1995). Second, the critical durations of [s] and [t] were set to an ambiguous value such that any durational effects must result from the speaking rate context. If the lack of rate effects on stress perception in eye tracking (Chapter 3) was indeed due to the presence of stress cues on the critical syllables, the use of ambiguous durations for boundary sounds should avoid this problem. Third, Chapter 5 addressed the amount and location of rate context that is necessary in order to show effects. Previous literature on rate effects suggested that rate context that is proximal to a target is more important than distal context (e.g., Summerfield, 1981). Several exceptions to this claim, however, have already been found (e.g., Kidd, 1989). Here, effects from proximal and distal contexts were compared. Fourth, by investigating such effects of proximal and distal contexts in both eye-tracking and categorization tasks, insights about the use of speaking rate information over time vs. at a later phase of processing were sought. These multiple questions allowed for a systematic investigation of how the speaking rate of a preceding context is assessed and consequently used in word recognition.

The first two experiments in Chapter 5 established that durational cues to word boundaries are perceived relative to speaking rate and that speaking rate information is taken into account immediately, that is, during the word recognition process. Speaking rate effects thus can be detected with eye tracking. The time course of rate effects was similar to the time course of the use of phoneme durations at the word boundary (Shatzman & McQueen, 2006). Three further experiments explored the role of the amount and location of the speaking rate context in eye tracking and categorization tasks. Proximal and distal

contexts were defined by dividing the carrier sentences into parts. The words immediately preceding the targets ('wel een' in 'Ze heeft wel eens (s)peer gezegd' - "She once said spear/pear" or 'nooit' in 'Ze heeft nooit (t)rap gezegd' - "She never said staircase/quick") were counted as proximal whereas everything preceding 'wel eens' and 'nooit' was distal. The longest proximal context at the slow speaking rate was 250 ms. In Experiment 3 proximal and distal contexts were set to opposite rates (e.g., proximal context was fast and distal context was slow). Experiment 3 showed that although listeners used the rate of the proximal context to interpret the boundary sounds, the rate of the distal context attenuated these effects (as compared to effects after a uniformly rate-manipulated sentence). These results were similar in eye-tracking and categorization tasks. The results thus seemed to confirm the claim that speaking rate information is mainly taken into account from a time window of approximately 300 ms preceding and following the target (Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). Proximal rate context is indeed more important than distal context (Summerfield, 1981). Nevertheless distal rate context did have an effect.

Two further experiments established conditions under which a more distal rate context can have effects (see also Kidd, 1989; Wayland et al., 1994). In Experiment 4 additional words were added to the distal context (i.e., 'Tijdens de lange vergadering heeft ze wel eens (s)peer gezegd' - "During the long meeting she once said spear/ pear"). In Experiment 5 the previously described sentences were used (e.g., 'Ze heeft wel eens (s)peer gezegd'). In both experiments the proximal context ('wel eens') was set to the same speaking rate as the target ('(s)peer') and the following context ('gezegd'). The proximal rate context was thus not informative for the disambiguation of the boundary sounds. With this lack of informativeness of immediate contexts, listeners used rate information from the distal contexts to interpret the boundary sounds. When comparing the different context conditions (i.e., long distal rate context in Experiment 4 vs. short distal rate context in Experiment 5), however, interesting task differences emerged. The amount of context played a larger role in categorization than in eye tracking. Whereas a long distal context affected word segmentation in both the online and offline tasks, a short distal context had an effect only in the offline task (categorization). Moreover, in categorization a long distal context had larger rate effects than a short distal context. The effect of length of the distal rate context was even sufficient to compensate for potential interference from the uninformative proximal context. When effects of the long distal context were compared to a shorter but proximally rate-manipulated context (i.e., the complete short sentence context in the first two experiments),

effects of the long distal context were stronger despite the distal location. No such differences were found in eye tracking.

This discrepancy between eye-tracking and categorization results supports the suggestion that processing time influences the use of speaking rate context. In categorization, listeners have more time to initiate their responses. In this situation they are able to process the whole rate context, hence effects of context length can be observed. In contrast, eye tracking measures listeners' continuous update of the available acoustic information during the word recognition process. Given this continuous processing of new information, rate context information is likely to be continuously updated as well. Listeners used this rate information to interpret the ambiguous boundary sounds. Therefore they gave more weight to proximal than distal contexts. Listeners thus use speaking rate information in online processing. The amount of rate context that is taken into account for processing, however, depends on the processing time available.

Task differences in the processing of speaking rate context were found for the perception of word boundaries (Chapter 5) as well as the perception of lexical stress patterns (eye tracking in Chapters 3 and categorization in Chapter 4). The results for stress and word-boundary perception in the eye-tracking experiments, however, were not the same. Whereas the word boundaries in Chapter 5 were interpreted relative to the speaking rate context online, no such effects were found for the perception of lexical stress (Chapter 3). One possible explanation for this discrepancy is the difference in the temporal characteristics of stress patterns and word boundaries. The processing of lexical stress patterns may require a focus on the alternating degrees of stress within a word. This focus on within-word durational relations could disguise effects from speaking rate context. Word boundaries, in contrast, are not distributed and thus can be affected by rate context. There are, however, at least two results that speak against this account. First, the eye tracking experiments reported in Chapter 2 and Chapter 3 demonstrated that stress information is taken into account as soon as it comes available. Perceiving the complete stress pattern of a word is thus not necessary to recognize words by means of stress cues. Second, the categorization task in Chapter 4 demonstrated that stress patterns are interpreted relative to speaking rate and that this influence of speaking rate appears to be incremental (rate-modulated stress cues were more important on initial than non-initial syllables). A more likely explanation for the difference in online rate effects between stress cues and cues to word boundary location is therefore the difference in the experimental stimuli. As mentioned before, in the experiments

in Chapter 3 durational stress cues were present on the initial two syllables of the target words. In contrast, the durations of the boundary sounds in Chapter 5 were set to a perceptually ambiguous value. Any measured effects thus had to result from the juncture phonemes' interpretation relative to speaking rate. A clear prediction follows: If in the materials used in Chapter 3 the durations of the first two syllables of the targets were set to durations that were ambiguous for stress, rate effects on stress processing should be observed in eye tracking.

These observations highlight an important difference between the eye-tracking and categorization tasks. In eye tracking the ambiguity of the critical durations appears to be a prerequisite for measuring time course effects of speaking rate. In contrast, in the categorization tasks in Chapter 4 (lexical stress) as well as in Chapter 5 (word boundaries) the perception of unambiguous sounds was also modulated by speaking rate context. Thus in order to gain a full picture of how acoustic cues are processed, a combination of methods that tap into different phases of word processing and different combinations of cues are necessary. A complete understanding of word recognition thus depends on comparing results from different stimuli and tasks.

Conclusions

This thesis explored how listeners process temporal information in the unfolding speech signal. Previous literature suggested that listeners incrementally use available acoustic information to recognize words optimally (Dahan, Magnuson, Tanenhaus, & Hogan, 2000; Norris & McQueen, 2008; Tanenhaus et al., 1995). The present investigation showed that listeners also use suprasegmental lexical stress cues, word boundary cues, and speaking rate context in an incrementally optimal fashion. Suprasegmental cues to lexical stress are used immediately to modulate lexical competition. If stress information comes available earlier than segmental information, words can be disambiguated by means of their stress patterns alone. This confirms that listeners use the same processing mechanism to evaluate segmental and suprasegmental information at a prelexical level of processing (Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Moreover, suprasegmental durational cues to lexical stress and to word boundary location are interpreted relative to speaking rate. As the speech signal unfolds, listeners not only update the current acoustic information to modulate lexical competition but also update their current estimates of speaking rate. This speaking rate estimate is then taken into account prelexically and incrementally to inform the evaluation of upcoming segmental and suprasegmental information. In addition, a

comparison of speaking rate effects from proximal and distal contexts in online and offline processing suggested that the amount of speaking rate context that is taken into account depends on the processing time available. During word recognition listeners give most weight to the rate context immediately preceding the durational information that is currently being evaluated. Thus, for a complete understanding word recognition it is necessary not only to investigate the use of different acoustic cues but also to compare tasks that explore different phases of the word recognition process. This thesis revealed important new information about the processing of the fine-temporal structure of spoken words. It thus can inform accounts of the inner workings of the word recognition process.

References

Allen, J., & Miller, J. L. (2001). Contextual influences on the internal structure of phonetic categories: A distinction between lexical status and speaking rate. *Perception & Psychophysics, 63*, 798-810.

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

Baayen, H., Davidson, D. J., & Bates, D. M. (2008). Mixed-effect modeling with crossed random effects for subjects and items. *Journal of Memory and Language, 59*, 390-412.

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

Barr, D. J. (2008). Analyzing 'visual world' eyetracking data using multilevel logistic regression. *Journal of Memory and Language, 59*, 457-474.

Bates, D. M., & Sarkar, D. (2007). *lme4: Linear mixed-effects models using S4 classes* (version 0.999375-27) [software application]. Retrieved from http://www.r-project.org

Boersma, P., & Weenink, D. (2007). PRAAT, a system for doing phonetics by computer (version 4.6.12) [computer program]. Retrieved from http://www.praat.org

Cho, T., McQueen, J. M., & Cox, E. A. (2007). Prosodically driven phonetic detail in speech processing: The case of domain-initial strengthening in English. *Journal of Phonetics, 35*, 210-243.

Cooper, N., Cutler, A., & Wales, R. (2002). Constraints of lexical stress on lexical access in English: Evidence from native and non-native listeners. *Language and Speech, 45*, 207-228.

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

Crystal, T. H., & House, A. S. (1982). Segmental durations in connected speech signals: Preliminary results. *Journal of the Acoustical Society of America, 72*, 705-716.

Crystal, T. H., & House, A. S. (1988). Segmental durations in connected-speech signals: Current results. *Journal of the Acoustical Society of America, 83*, 1553-1573.

Cutler, A. (1986). Forbear is a homophone: Lexical prosody does not constrain lexical access. *Language and Speech, 29*, 201-220.

Cutler, A., & Chen, H.-C. (1997). Lexical tone in Cantonese spoken word-processing. *Perception & Psychophysics, 59*, 165-179.

Cutler, A., & van Donselaar, W. (2001). Voornaam is not (really) a homophone: Lexical prosody and lexical access in Dutch. *Language and Speech, 44*, 171-195.

Cutler, A., & Pasveer, D. (2006). Explaining cross-linguistic differences in effects of lexial stress on spoken-word recognition. In R. Hoffmann & H. Mixdorff (Eds.), *Proceedings of Speech Prosody 2006* (pp. 237-240). Dresden: TUD Press.

Cutler, A., Wales, R., Cooper, N., & Janssen, J. (2007). Dutch listeners' use of suprasegmental cues to English stress. In J. Trouvain & W. J. Barry (Eds.), *Proceedings of the XVI International Congress of Phonetic Sciences* (pp. 1913-1916). Dudweiler: Pirrot.

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

Dilley, L. C., & McAuley, D. J. (2008). Distal prosodic context affects word segmentation and lexical processing. *Journal of Memory and Language, 59*, 294-311.

Fry, D.B. (1955). Duration and intensity as physical correlates of linguistic stress. *Journal of the Acoustical Society of America, 27*, 765-768.

Gay, T. (1978). Effect of speaking rate on vowel formant movements. *Journal of the Acoustical Society of America, 63*, 223-230.

Gordon, P. C. (1988). Induction of rate-dependent processing by coarse-grained aspects of speech. *Perception & Psychophysics, 43*, 137-146.

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.

Green, K. P., Stevens, E. B., & Kuhl, P. K. (1994). Talker continuity and the use of rate information during phonetic perception. *Perception & Psychophysics, 55*, 249-260.

Green, K. P., Tomiak, G. R., & Kuhl, P. K. (1997). The encoding of rate and talker information during phonetic perception. *Perception & Psychophysics, 59*, 675-692.

Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman & J. P. Thomas (Eds.), *Handbook of perception and human performance* (pp. 10-11--10-112). New York: Wiley.

Huettig, F., & McQueen, J. M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 57*, 460-482.

Janse, E. (2003). Word perception in fast speech: artificially time-compressed vs. naturally produced fast speech. *Speech Communication, 42*, 155-173.

Kidd, G. R. (1989). Articulatory-rate context effects in phoneme identification. *Journal of Experimental Psychology: Human Perception and Performance, 15*, 736-748.

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.

Matin, E., Shao, K. C., & Boff, K. R. (1993). Saccadic overhead: Information-processing time with and without saccadic overhead. *Perception & Psychophysics, 53*, 372-380.

Mattys, S. L. (2000). The perception of primary and secondary stress in English. *Perception and Psychophysics, 62*, 253-265.

Mattys, S. L., White, L., & Melhorn, J. F. (2005). Integration of multiple speech segmentation cues: A hierarchical framework. *Journal of Experimental Psychology: General, 134*, 477-500.

McQueen, J. M., & Viebahn, M. (2007). Tracking recognition of spoken words by tracking looks to printed words. *The Quarterly Journal of Experimental Psychology, 60*, 661-671.

Miller, J. L. (1981). Effects of speaking rate on segmental distinctions. In P. D. Eimas & J. L. Miller (Eds.), *Perspectives on the Study of Speech.* (pp. 39-74). Hillsdale, NJ: Erlbaum Associates.

Miller, J. L. (1987). Rate-dependent processing in speech perception. In A. W. Ellis (Ed.), *Progress in the Psychology of Language* (Vol. 3, pp. 119-157). London: Erlbaum Associates.

Miller, J. L., & Dexter, E. R. (1988). Effects of speaking rate and lexical status on phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 14,* 369-378.

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.

Miller, J. L., & Wayland, S. C. (1993). Limits on the limitations of context-conditioned effects in the perception of [b] and [w]. *Perception & Psychophysics, 54,* 205-210.

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

Newman, R. S., & Sawusch, J. R. (1996). Perceptual normalization for speaking rate: Effects of temporal distance. *Perception & Psychophysics, 58,* 540-560.

Newman, R. S., & Sawusch, J. R. (2009). Perceptual normalization for speaking rate III: Effects of the rate of one voice on perception of another. *Journal of Phonetics, 37,* 46-65.

Nooteboom, S. G. (1972). *Production and perception of vowel duration, a study of durational properties of vowels in Dutch.* Doctoral Dissertation, Utrecht University.

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

Norris, D., & McQueen, J. M. (2008). Shortlist B: A Bayesian model of continuous speech recognition. *Psychological Review, 115,* 357-395.

Port, R. (1979). The influence of tempo on stop closure duration as a cue for voicing and place. *Journal of Phonetics, 7,* 45-56.

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

Reinisch, E., Jesse, A., & McQueen, J. M. (2008a). Lexical stress information modulates the time-course of spoken-word recognition. *Proceedings of Acoustics'08 (CD-ROM),* (pp. 3183-3188). Paris: Société Française d'Acoustique.

Reinisch, E., Jesse, A., & McQueen, J. M. (2008b). The strength of lexical competition depends on the presence of first-syllable stress. *Proceedings of Interspeech 2008 (CD-ROM).* (p. 1954).

Reinisch, E., Jesse, A., & McQueen, J. M. (2010). Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately. *The Quarterly Journal of Experimental Psychology, 63,* 772-783.

Reinisch, E., Jesse, A., & McQueen, J. M. (in press). Speaking rate affects the perception of duration as a suprasegmental lexical-stress cue. *Language and Speech.*

Repp, B. H., Liberman, A. M., Eccardt, T., & Pesetsky, D. (1978). Perceptual integration of acoustic cues for stop, fricative, and affricate manner. *Journal of Experimental Psychology: Human Perception and Performance, 4,* 621-637.

Rietveld, T., Kerkhoff, J., & Gussenhoven, C. (2004). Word prosodic structure and vowel duration in Dutch. *Journal of Phonetics, 32,* 349-371.

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

Sawusch, J. R., & Newman, R. S. (2000). Perceptual normalization for speaking rate II: Effects of signal discontinuities. *Perception & Psychophysics, 62,* 285-300.

Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics, 68,* 1-16.

Slootweg, A. (1988). Metrical prominence and syllable duration. In P. Coopmans, & A. Hulk (Eds.), *Linguistics in the Netherlands 1988* (pp. 139–148). Dordrecht: Fortis Publications.

Sluijter, A. M. C., & van Heuven, V. J. (1995). Effects of focus distribution, pitch accent and lexical stress on the temporal organization of syllables in Dutch. *Phonetica, 52,* 71-89.

Sluijter, A. M. C., & van Heuven, V. J. (1996). Spectral balance as an acoustic correlate of linguistic stress. *Journal of the Acoustical Society of America, 100,* 2471-2485.

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

Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language, 45,* 412-432.

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

Summerfield, Q. (1981). Articulatory rate and perceptual constancy in phonetic perception. *Journal of Experimental Psychology: Human Perception and Performance, 7,* 1074-1095.

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

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

van Alphen, P. M., & Smits, R. (2004). Acoustical and perceptual analysis of the voicing distinction in Dutch initial plosives: the role of prevoicing. *Journal of Phonetics, 32,* 455-491.

van Donselaar, W., Koster, M., & Cutler, A. (2005). Exploring the role of lexical stress in lexical recognition. *The Quarterly Journal of Experimental Psychology, 58A,* 251-273.

van Heuven, V. J. (1988). Effects of stress and accent on the human recognition of word fragments in spoken context: gating and shadowing. In W. A. Ainsworth, & J. N. Holmes (Eds.), *Proceedings of the 7th FASE/Speech-88 Symposium* (pp. 811-818). Edinburgh, UK: The Institute of Acoustics.

van Heuven, V. J., & Hagman, P. (1988). Lexical statistics and spoken word recognition in Dutch. In P. Coopmans & A. Hulk (Eds.), *Linguistics in the Netherlands 1988* (pp. 59-68). Dordrecht: Fortis.

van Heuven, V. J., & Menert, L. (1996). Why stress position bias? *Journal of the Acoustical Society of America, 100,* 2439-2451.

van Leyden, K., & van Heuven, V. J. (1996). Lexical stress and spoken word recognition: Dutch vs. English. In M. Den Dikken & C. Cremers (Eds.), *Linguistics in the Netherlands 1996* (pp. 159-170). Amsterdam: John Benjamins.

Wayland, S. C., Miller, J. L., & Volaitis, L. E. (1994). The influence of sentential speaking rate on the internal structure of phonetic categories. *Journal of the Acoustical Society of America, 95,* 2694-2701.

Wightman, C. W., Shattuck-Hufnagel, S., Ostendorf, M., & Price, P. J. (1992). Segmental durations in the vicinity of prosodic phrase boundaries. *Journal of the Acoustical Society of America, 91,* 1707–1717.

Samenvatting en conclusies

Het doel van dit proefschrift was te onderzoeken hoe luisteraars gebruik kunnen maken van temporele informatie voor de herkenning van woorden in het continue spraaksignaal van hun moedertaal. Zoals een aantal keren besproken in dit proefschrift, breidt het spraaksignaal zich niet met een gelijkmatige snelheid uit. Een zin kan sneller of langzamer worden uitgesproken. Bovendien is de lengte van klanken niet alleen afhankelijk van hun articulatorische eigenschappen en spraaksnelheid, maar ook van structurele invloeden zoals hun positie in de prosodische structuur van een woord en de locatie van dit woord in een zin. Dit proefschrift onderzoekt hoe luisteraars met dit soort temporele variatie omgaan, dat wil zeggen, of en hoe luisteraars temporele variatie gebruiken tijdens het proces van woordherkenning.

Om gesproken taal te decoderen maken luisteraars meteen gebruik van de beschikbare klankinformatie om voortdurend hun hypotheses te vernieuwen over het woord dat mogelijk wordt gehoord. In dit proefschrift wordt onderzocht of luisteraars ook suprasegmentele informatie gebruiken om woorden te herkennen, in bijzonder, in hoeverre de perceptie van suprasegmentele temporele aanwijzingen voor de locatie van klemtoon en van woordgrenzen rekening houdt met de spraaksnelheid van de voorafgaande context. De focus van dit onderzoek was de verwerking van dit soort aanwijzingen terwijl het spraaksignaal zich uitbreidt ("online" verwerking). Resultaten van online verwerking, gemeten met oogbewegingsexperimenten, worden vervolgens vergeleken met resultaten van taken die beïnvloed worden door post-perceptuele fases van het spraakverwerkingsproces. Deze vergelijking was een poging om het woordherkenningsproces in actie te verkennen maar tegelijkertijd de beperkingen van dit proces te bepalen. Verder wordt onderzocht hoe spraaksnelheid wordt geëvalueerd tijdens het spraakverwerkingsproces versus als er meer verwerkingstijd beschikbaar is. Dit wordt bereikt door de invloeden van direct voorafgaande context te vergelijken met de invloeden van vroegere contexten.

Oogbewegingsexperimenten met het visuele wereld paradigma ('visual world paradigm') zijn een voortreffelijke methode om het verloop van woordherkenning te onderzoeken. Eerder onderzoek met dit paradigma toonde aan dat luisteraars meteen gebruik kunnen maken van klankinformatie, en dat fonetisch detail kan worden gebruikt om woordgrenzen te herkennen (Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). De oogbewegingsexperimenten in dit proefschrift onderzoeken wanneer luisteraars suprasegmentele aanwijzingen voor klemtoon en spraaksnelheid gebruiken. In alle experimenten keken luisteraars naar geschreven woorden op het computerscherm (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Door het gebruik van geschreven woorden in plaats van plaatjes was het mogelijk een groter aantal doelwoorden en rivaalwoorden ("competitors") te gebruiken, omdat de woorden niet verbeeld hoefden te worden. Bovendien kan met geschreven woorden de verwerking van fonologische informatie worden bestudeerd (Huettig & McQueen, 2007).

Hoofdstuk 2: Vroeg gebruik van klemtooninformatie

Het experiment in hoofdstuk 2 beschrijft onderzoek naar het tijdelijke verloop van het woordherkenningsproces als Nederlandse luisteraars het klemtoonpatroon van een woord gebruiken om woorden te herkennen. Klemtoonpatronen geven temporele informatie op een grotere tijdscala weer dan segmentale informatie, omdat klemtoonpatronen over de lettergrepen van een woord zijn verspreid. Daarom vereist de verwerking van klemtoonpatronen tijdens woordherkenning de herkenning van afwisselende sterktes van klemtonen. Anders als in het Engels, waar niet-beklemtoonde klinkers worden gereduceerd, wordt een beklemtoonde lettergreep in het Nederlands meestal alleen suprasegmenteel gemarkeerd (met duur, toonhoogte, luidheid, en de verdeling van energie in het spectrum). Met de uitvoering van de experimenten in het Nederlands was het dus mogelijk effecten van suprasegmentele klemtooninformatie zonder de invloed van effecten door segmentele indicatie van klemtoon te onderzoeken. Eerder onderzoek veronderstelde dat het gebruik van klemtoon in het Nederlands een voordeel is voor de herkenning van woorden (Cutler & Pasveer, 2006; van Heuven & Hagman, 1988). Het gebruik van klemtooninformatie vermindert het aantal woorden dat ook deel uitmaakt van een langer woord (bijvoorbeeld 'ham' in 'hamster'; Cutler & Pasveer, 2006). Verder kunnen woorden met behulp van klemtooninformatie sneller van andere woorden worden onderscheiden (van Heuven & Hagman, 1988). Nederlandse luisteraars gebruiken inderdaad suprasegmentele klemtooninformatie in de herkenning van woorden (e.g., Cooper, Cutler, & Wales, 2002;

Cutler & van Donselaar, 2001; van Donselaar, Koster, & Cutler, 2005; van Heuven, 1988; van Leyden & van Heuven, 1996). Een voorbeeld is het "fragment priming" experiment dat wordt beschreven in van Donselaar et al. (2005). "Fragment priming" houdt in dat luisteraars eerst een fragment van een woord horen en vervolgens een geschreven woord op een computerscherm zien. De proefpersoon moet beslissen of het geschreven woord een echt woord is of een non-woord. Het experiment liet zien dat een met het doelwoord segmenteel overlappende prime de beslissing versnelt als deze ook qua klemtoonpatroon overeenkomt met het geschreven doelwoord, zelfs als de prime maar een lettergreep lang was (bijvoorbeeld, de prime 'OC-' helpt de herkenning van 'OCtopus'; hoofdletters tonen beklemtoonde lettergrepen). Primes met klemtoonpatronen die niet overeenkwamen met het klemtoonpatroon van het doelwoord verstoorden de herkenning van het doelwoord. Dit was het geval als hun lengte twee lettergrepen omvatte ('okTO-' 'OCtopus'). Niet overeenkomende primes, die maar één lettergreep omvatten, verstoorden de herkenning van het doelwoord niet, maar ze hielpen de herkenning ook niet. Eerder onderzoek stelde dus voor dat de hoeveelheid aan klemtooninformatie die luisteraars beschikbaar hebben (een vs. twee lettergrepen) de woordherkenning beïnvloedt. Geen van deze eerdere experimenten kon echter aantonen wanneer precies tijdens het woordherkenningsproces klemtooninformatie wordt gebruikt.

Daarom wordt in hoofdstuk 2 een oogbewegingsexperiment beschreven. Deze methode maakt het mogelijk het tijdelijke verloop van lexicale competitie te onderzoeken, terwijl in het begin de competitie van alleen klemtooninformatie kan worden gemoduleerd. Nederlandse doel- en rivaalwoorden die in hun eerste twee lettergrepen dezelfde klanken maar verschillende klemtoonpatronen hebben werden gekozen. Er waren twee types woordparen, ten eerste, woordparen met de primaire klemtoon op hun eerste vs. tweede lettergreep ('OCtopus' - 'okTOber', 1-2 contrast) en, ten tweede, woordparen met de primaire klemtoon op hun eerste vs. derde lettergreep ('DIameter' – 'diaMANT', 1-3 contrast). Nederlandse woorden met een primaire klemtoon op hun derde lettergreep hebben een secundaire klemtoon op hun eerste lettergreep. Elk woord hoorde dus bij een van vier condities die door de combinatie van klemtooncontrast (1-2 vs. 1-3 contrast) en klemtoonlocatie (initieel vs. niet initieel) resulteerden.

De resultaten toonden aan dat luisteraars optimaal gebruik kunnen maken van suprasegmentele klemtooninformatie en dat dit plaatsvindt zodra de klemtooninformatie beschikbaar wordt. Luisteraars herkenden woorden gebaseerd op hun klemtoonpatroon voordat segmentele klankinformatie een eenduidige herkenning toeliet. Deze resultaten voor drie van de vier condities gevonden. Alleen woorden met een primaire

klemtoon op de derde lettergreep waren moeilijk te onderscheiden op basis van alleen hun klemtoonpatroon. Omdat deze woorden een secundaire klemtoon op hun eerste lettergreep hebben en beide woorden van de 1-3 contrast een onbeklemtoonde tweede lettergreep hebben, zijn hun eerste twee lettergrepen akoestisch bijna gelijk. Daarom wordt verwacht dat woorden met primaire klemtoon op hun derde lettergreep het moeilijkst zijn te herkennen op basis van alleen hun klemtoonpatroon. Verder, lijken initieel beklemtoonde woorden ('OCtopus') sterker voor herkenning te vechten dan niet initieel beklemtoonde woorden ('okTOber') – onafhankelijk van het klemtoonpatroon van het doelwoord. Deze resultaten laten zien dat klemtooninformatie wordt gebruikt tijdens het woordherkenningsproces en dat klemtooncontrast (1-2 vs. 1-3 contrast) en klemtoonlocatie (initieel vs. niet initieel) een invloed hebben hoe sterk een woord voor herkenning vecht.

Deze resultaten dragen ook bij aan een andere discussie over klemtoonperceptie. Eerder onderzoek vond dat luisteraars een voorkeur hebben voor woorden met de klemtoon in woord-initiële positie (van Leyden & van Heuven, 1996). Deze voorkeur was meestal toegeschreven aan het feit dat in het Nederlands (en trouwens ook in het Engels) meer woorden een klemtoon op de eerste lettergreep hebben dan op een niet-initiële lettergreep (van Heuven & Hagman, 1988). In het experiment van hoofdstuk 2 worden vroege verschillen in de sterkte van competitie tussen doelwoorden met initiële en niet-initiële klemtoon gevonden. Woorden met klemtoon op de eerste lettergreep vochten sterker om herkenning dan woorden met klemtoon op een andere lettergreep. Dit wijst erop dat het vroegere voorstel van een voorkeur naar de perceptie van woord-initiële klemtoon niet alleen afkomstig is van de statistische distributie van klemtoonpatronen. Vaak wordt de sterkere competitie van initieel beklemtoonde woorden gedreven door het akoestisch signaal. De aanwezigheid van klemtoon is makkelijker te herkennen dan de afwezigheid van klemtoon. Als een lettergreep wordt waargenomen als niet-beklemtoond, is het mogelijk dat luisteraars onzeker zijn of dat er geen aanwijzing is voor klemtoon, of dat de aanwijzing is gemist. Hoewel de vroeger aangetoonde statistische voorkeur tijdens een latere fase van het woordherkenningsproces wel een rol kan spelen, blijkt er toch een vroege, door het akoestisch signaal gedreven effect aanwezig te zijn, die aanleiding geeft voor luisteraars om de aanwezigheid van klemtoon meer gewicht toe te kennen dan de afwezigheid van klemtoon.

Met een correlatieanalyse wordt verder onderzoekt welke akoestische aanwijzingen voor klemtoon door luisteraars worden gebruikt. De enige aanwijzing die de oogbewegingen naar het doelwoord kon voorspellen was tijdsduur. Een groter verschil tussen de duur van initieel en niet initieel beklemtoonde klinkers correleerde met een groter

verschil in fixaties op initieel en niet initieel beklemtoonde woorden. Dit resultaat sluit aan bij resultaten van Nooteboom (1972) en Sluijter en van Heuven (1995, 1996) die veronderstelden dat in het Nederlands duur de meest consistente aanwijzing voor klemtoon is. Op basis van deze vondsten rees de vraag of de waarneming van klemtoon op een bepaalde lettergreep afhankelijk is van de spraaksnelheid in een voorafgaande context. Deze vraag wordt in de hoofdstukken 3 en 4 onderzocht. Hoofdstuk 2 toonde aan dat temporele klemtooninformatie in een woord zo vroeg mogelijk wordt gebruikt om lexicale competitie te moduleren.

Hoofdstuk 3: De waarneming van klemtoon in relatie met spraaksnelheid

De interpretatie van temporele informatie is fundamenteel afhankelijk van de spraaksnelheid van een zin. Eerdere literatuur toonde aan dat de interpretatie van duurinformatie over klanken beïnvloed wordt door de spraaksnelheid van de context (bijv. Allen & Miller, 2001; Miller, 1981; 1987; Miller & Dexter, 1988; Miller & Liberman, 1979; Miller & Wayland, 1993; Wayland, Miller, & Volaitis, 1994). Als bijvoorbeeld het Engelse woord 'bin' ("bak") in een snelle context met hoge spraaksnelheid wordt geplaatst, zullen luisteraars vermelden dat ze 'pin' ("speld") hebben gehoord. In vergelijking met de korte klanken bij een hoge spraaksnelheid lijkt de "voice onset time" (de tijd tussen het weer vrijgeven van de luchtstroom naar een constrictie in een plofklank en het begin van stembandvibratie in de volgende klinker) van de /b/ langer en lijkt deze dus op een /p/. De meest gebruikte taken voor dit soort onderzoek waren classificatietaken en de beoordeling hoe goed een klank bij een bepaalde categorie past. Beide taken meten perceptuele en post-perceptuele fases van het woordherkenningsproces. Desondanks stelden enkele studies voor dat effecten van spraaksnelheid tijdens een vroege fase van klankperceptie optreden (Miller & Dexter, 1988; Newman & Sawusch, 2009; Sawusch & Newman, 2000). Luisteraars gebruikten informatie over spraaksnelheid van andere sprekers dan de doelspreker (Green, Stevens, & Kuhl, 1994; Green, Tomiak, & Kuhl, 1997; Newman & Sawusch, 2009; Sawusch & Newman, 2000). Dit was het geval zelfs als deze andere spreker zich op een andere locatie bevond als de doelspreker (Newman & Sawusch, 2009). Informatie over spraaksnelheid wordt dus verwerkt voordat andere processen zoals perceptuele samenstelling of onderscheiding van klankstromen plaatsvinden. Omdat duur de meest consistente aanwijzing is voor de aanwezigheid van klemtoon, en omdat klemtooninformatie tijdens een vroege fase van het woordherkenningsproces wordt verwerkt (zoals getoond in hoofdstuk 2), wordt verwacht dat spraaksnelheid een invloed zal hebben op de waarneming

van klemtoonsterkte. Lexicale competitie tussen initieel en niet initieel beklemtoonde woorden zal worden gemoduleerd door de spraaksnelheid van een voorafgaande context. Na een langzame context zal de initiële lettergreep van een woord korter klinken en daarom als minder beklemtoond worden waargenomen dan na een snelle context. Dit is het onderwerp van hoofdstuk 3.

In de experimenten in hoofdstuk 3 wordt dezelfde taak en hetzelfde materiaal gebruikt als in het experiment in hoofdstuk 2. In het eerste experiment werd alleen de voorafgaande context gemanipuleerd. De doelwoorden bevatten alle aanwijzingen voor klemtoon (duur, toonhoogte, luidheid). In dit eerste experiment werden er echter geen effecten van spraaksnelheid gevonden. Daarom werd in een tweede experiment onderzocht of effecten mogelijk door de aanwezigheid van andere aanwijzingen voor klemtoon worden versluierd (andere aanwijzingen dan duur zoals toonhoogte en luidheid). Daarom werden in het tweede experiment deze aanwijzingen geneutraliseerd. In elk woordpaar werden aanwijzingen van toonhoogte en luidheid op de eerste twee lettergrepen op hun gemiddelde waarden gezet. Het tweede experiment liet zien dat met alleen duur als aanwijzing voor klemtoon alle woorden sterker voor herkenning vochten dan als er meerdere types aanwijzingen aanwezig waren (zoals in het eerste experiment). Spraaksnelheid had echter weer geen invloed op de sterkte van de competitie. Een snelle spraaksnelheid leidde in het algemeen tot snellere herkenning van het doelwoord, maar dit was onafhankelijk van de klemtoonlocatie. Desondanks was duur een voldoende aanwijzing om woorden te herkennen voordat segmentele informatie beschikbaar kwam. De experimenten in hoofdstuk 3 repliceerden de vondst van het experiment in hoofdstuk 2. Luisteraars gebruiken klemtooninformatie onmiddellijk ter herkenning van woorden.

Het feit dat duur een voldoende aanwijzing was voor luisteraars om woorden middels hun klemtoonpatroon te herkennen kan echter de oorzaak zijn voor het uitblijven van spraaksnelheidseffecten. Het zou kunnen dat spraaksnelheid niet genoeg extra informatie kan geven bovenop het effect van duurinformatie op de doelwoorden. De resultaten in hoofdstuk 3 stellen dus voor dat spraaksnelheid het online woordherkenningsproces zoals het met oogbewegingsexperimenten wordt gemeten niet beïnvloed.

Hoofdstuk 4: Spraaksnelheid beïnvloedt het gebruik van aanwijzingen voor klemtoon

Om de relatie tussen spraaksnelheid en klemtoonperceptie verder te onderzoeken wordt in een serie van experimenten in hoofdstuk 4 met een andere taak getest of effecten van spraaksnelheid afhankelijk zijn van de aanwezigheid van aanwijzingen voor klemtoon op initiële en niet-initiële lettergrepen van doelwoorden. In een fragment classificatietaak moesten luisteraars besluiten van welk van twee woorden een woordfragment was afgeknipt. Het fragment wordt aan het einde van een snelle of langzame zin gepresenteerd. Voor de manipulatie wordt een deel van de woorden uit de eerder beschreven experimenten geselecteerd. Een woordpaar hoorde bij het 1-2 contrast ('Alibi'–'aLInea', fragment 'ali') of bij het 1-3 contrast ('CAvia'–'kaviAAR', fragment 'kavi'). De duur van de eerste of tweede klinker van de fragmenten was gemanipuleerd om een duurcontinuüm te genereren. De andere klinker wordt naar de gemiddelde duur van beklemtoonde en niet-beklemtoonde klinkers in deze lettergreep aangepast (de lengte van die klinker was dus ambigue). Op die manier was het mogelijk verdere vragen over effecten van spraaksnelheid op initiële en niet-initiële lettergrepen te behandelen.

Hier werden sterke effecten van spraaksnelheid op de perceptie van klemtoonpatronen gevonden. Spraaksnelheid beïnvloedde de waarneming van klemtoon op de eerste en tweede lettergreep van woorden in beide klemtooncontrasten (1-2 en 1-3 contrast). De effecten van spraaksnelheid waren verder niet afhankelijk van de perceptie van aanwijzingen voor klemtoon op deze lettergrepen zelf. Hoewel de perceptie van duur op de tweede lettergrepen wel beïnvloed wordt door spraaksnelheid, wordt de duur van het continuüm niet gebruikt in de beoordeling van klemtoonsterkte. Deze uitkomst ondersteunt de vondsten van hoofdstuk 2 en 3 dat klemtooninformatie gradueel wordt gebruikt. Bovendien waren aanwijzingen voor klemtoon (relatief geïnterpreteerd ten opzichte van spraaksnelheid) belangrijker op initiële dan op niet-initiële lettergrepen. Samen laten de sterkere spraaksnelheidseffecten op initiële lettergrepen en de zwakke invloed van duurinformatie op de tweede lettergreep zien, dat spraaksnelheid op een graduele manier wordt gebruikt om duurinformatie naar klemtoon te interpreteren.

De resultaten van een gedragstaak wijzen erop dat de waarneming van duurinformatie over klemtoonlocatie afhankelijk is van de spraaksnelheid van een voorafgaande context (hoofdstuk 4). Hoofdstuk 3 maakte gebruik van oogbewegingmeting en vond geen dergelijke effecten. Tijdens het woordherkenningsproces vernieuwen luisteraars voortdurend de beschikbare akoestische informatie om woorden in het mentale

woordenboek op te zoeken ("lexical access"). Voor klemtoonperceptie is het echter mogelijk dat de relatie van aanwijzingen voor klemtoon tussen de lettergrepen in een woord sterker is dan de relatie tussen de klemtooninformatie op de eerste lettergreep en de voorafgaande context. Op die manier zou de aanwezigheid van aanwijzingen voor klemtoon (zelfs als het maar duur was) op de doelwoorden in de oogbewegingstaak alle effecten van spraaksnelheid versluierd kunnen hebben. Als luisteraars echter meer tijd hebben om de stimuli te verwerken (en dus ook de voorafgaande context) – zoals in de classificatietaak – zullen ze wel spraaksnelheid kunnen gebruiken voor de interpretatie van klemtoon, zelfs als er klemtooninformatie op de lettergrepen in de doelwoorden aanwezig is. Niettemin wijzen de gevonden spraaksnelheidseffecten bij de verschillende lettergrepen in de classificatietaak erop dat spraaksnelheid gradueel wordt gebruikt, zoals het wordt verwacht tijdens vroege fases van het woordherkenningsproces. Daarom wordt in hoofdstuk 5 verder onderzocht wanneer tijdens het woordherkenningsproces informatie over spraaksnelheid kan worden gebruikt. Hoofdstuk 5 heeft verder belangstelling voor de vraag of veranderen van de experimentele taak verschillen in de verwerking van spraaksnelheid kan veroorzaken, een vraag die uit het verschil in resultaten van de hoofdstukken 3 en 4 naar voren kwam.

Hoofdstuk 5: Spraaksnelheid beïnvloed woordsegmentatie

De serie van experimenten die in hoofdstuk 5 wordt gepresenteerd, breidt de onderwerpen van de eerdere hoofdstukken op verschillende manieren uit. Ten eerste wordt onderzocht of spraaksnelheidseffecten ook op andere types suprasegmentele duurinformatie kunnen worden toegepast, namelijk op duur als aanwijzing voor de locatie van woordgrenzen. Ambigue samenvoegingen van woorden zoals 'wel eens (s)peer' en 'nooit (t)rap' worden gebruikt om de perceptie van de laatste woorden van deze twee-woord frases te toetsen. Beginnen ze met een [s] of [t] of niet? Langere klanken aan de woordgrenzen leiden tot een waarneming van deze grensklanken als woord-initieel (Gow & Gordon, 1995; Klatt, 1976; Quené, 1992; Repp, Liberman, Eccardt, & Pesetsky, 1978; Salverda et al., 2003; Shatzman & McQueen, 2006; Spinelli, McQueen, & Cutler, 2003; Tabossi, Burani, & Scott, 1995). Ten tweede wordt de kritische duur van [s] en [t] aangepast tot een ambigue waarde. Op deze manier moeten alle effecten van duur afkomstig zijn van de spraaksnelheid in de context. Als de afwezigheid van spraaksnelheidseffecten in de oogbewegingsexperimenten in hoofdstuk 3 inderdaad het gevolg was van de aanwezigheid van klemtooninformatie op de kritische lettergrepen, zal het gebruik van ambigue klanken dit probleem oplossen. Ten derde onderzoekt hoofdstuk 5 de vraag naar de lengte en locatie

van de context die nodig is om effecten van spraaksnelheidsinformatie te tonen. Eerdere literatuur veronderstelde dat spraaksnelheidsinformatie nabij een doelklank grotere invloed heeft op de perceptie dan verafgelegen informatie (Summerfield, 1981). Een aantal uitzonderingen hierop zijn echter al gevonden (Kidd, 1989). In hoofdstuk 5 worden effecten van nabije en verafgelegen contexten van spraaksnelheidsinformatie vergeleken. Ten vierde zal door de vergelijking van effecten van nabije en verafgelegen contexten in oogbewegingsexperimenten en classificatietaken worden uitgezocht hoe spraaksnelheid wordt verwerkt tijdens de uitbreiding van het spraaksignaal versus tijdens een latere fase van het verwerkingsproces. Deze vragen hebben geleid tot een systematisch onderzoek over de berekening van spraaksnelheid en hoe deze informatie vervolgens gebruikt wordt om woorden te herkennen.

De eerste twee experimenten in hoofdstuk 5 hebben aangetoond dat duur als aanwijzing voor de locatie van woordgrenzen relatief ten opzichte van spraaksnelheid wordt waargenomen en dat spraaksnelheidsinformatie meteen gebruikt wordt tijdens het woordherkenningsproces. Effecten van spraaksnelheid kunnen dus inderdaad met oogbewegingsexperimenten worden gemeten. Het temporele verloop van deze effecten lijkt op het gebruik van duurinformatie op de grensklanken zelf (Shatzman & McQueen, 2006). Drie aanvullende experimenten onderzochten de rol van de lengte en locatie van de context in oogbewegingstaken en classificatietaken. De locatie van de context wordt bepaald door de contextzinnen op te splitsen in een verafgelegen en nabije context. De woorden die het doelwoord direct voorafgaan ('wel eens' in 'Ze heeft wel eens (s)peer gezegd' en 'nooit' in 'Ze heeft nooit (t)rap gezegd') worden als nabij geclassificeerd, terwijl alles voor 'wel eens' en 'nooit' als "verafgelegen" wordt geclassificeerd. In experiment 3 werden de nabije en verafgelegen contexten naar hun tegengestelde spraaksnelheid aangepast (de nabije context was snel en de verafgelegen context was langzaam, en andersom). Experiment 3 toonde aan dat luisteraars de nabije context voor de interpretatie van de grensklanken gebruikten, maar dat dit effect door de tegengestelde spraaksnelheid van de verafgelegen context wordt verminderd (wanneer de resultaten worden vergeleken met een zin met maar één spraaksnelheid). Deze resultaten waren vergelijkbaar in de oogbewegingstaken en classificatietaken. De resultaten bleken de suggestie te ondersteunen dat spraaksnelheid binnen een tijd van 300 ms voor en na een doelklank wordt berekend (Newman & Sawusch, 1996; Sawusch & Newman, 2000; Summerfield, 1981). De nabije context is inderdaad belangrijker dan de verafgelegen context (Summerfiel 1981). Desalniettemin was er ook een effect van verafgelegen context.

Twee verdere experimenten zochten uit onder welke omstandigheden een verafgelegen context effecten kan tonen. In experiment 4 werd de verafgelegen context verlengd door meerdere woorden aan het begin toe te voegen ('Tijdens de lange vergadering heeft ze wel eens (s)peer gezegd'). In experiment 5 werden de eerder beschreven zinnen gebruikt ('Ze heeft wel eens (s)peer gezegd'). In beide experimenten werd de nabije context ('wel eens') naar een neutrale spraaksnelheid aangepast. Dit was dezelfde spraaksnelheid waarmee ook het doelwoord ('(s)peer') en de volgende context ('gezegd') waren uitgesproken. De nabije context was dus niet informatief om de ambigue grensklanken te interpreteren. Zonder deze informatie van de nabije context bleken luisteraars de verafgelegen context te gebruiken om de grensklanken te interpreteren. Als de effecten van de verschillende condities worden vergeleken (de lange verafgelegen context in experiment 4 vs. de korte verafgelegen context in experiment 5), worden echter interessante verschillen tussen de experimentele taken gevonden. De lengte van de context was belangrijker in de classificatietaak dan in de oogbewegingstaak. De lange verafgelegen context beïnvloedde woordsegmentatie in beide taken. Een korte context werd echter alleen in de classificatietaak gebruikt. Bovendien had in de classificatietaak de lang verafgelegen context grotere invloed dan een korte verafgelegen context. Het effect van de lengte van de verafgelegen context kon zelfs voor de verafgelegen locatie compenseren. Als de effecten van de lange verafgelegen context worden vergeleken met een uniform gemanipuleerde onmiddellijke maar kortere context (de hele voorafgaande context in de experimenten 1 en 2) waren effecten van de lange verafgelegen context sterker dan die van de onmiddellijke maar kortere context. Een dergelijke verschil wordt in de oogbewegingstaken niet gevonden.

Dit verschil tussen resultaten van oogbewegingsexperimenten en classificatietaken ondersteunt de hypothese dat de tijd die luisteraars beschikbaar hebben om de context te verwerken invloed heeft op de gemeten effecten. In classificatietaken hebben luisteraars meer tijd om hun beslissing te nemen. Daarom kunnen ze de hele context verwerken en worden er effecten van contextlengte gevonden. Oogbewegingsexperimenten meten echter het continue gebruik van nieuw beschikbare informatie tijdens het woordherkenningsproces. Door deze continue verwerking van nieuwe informatie is het waarschijnlijk dat informatie over de spraaksnelheid ook continu wordt vernieuwd. Daarom besteden luisteraars meer aandacht aan de nabije dan de verafgelegen context in de interpretatie van de grensklanken. Luisteraars gebruiken informatie over spraaksnelheid online, tijdens het woordherkenningsproces. Hoeveel contextinformatie kan worden gebruikt is echter afhankelijk van de beschikbare verwerkingstijd.

Verschillen in de verwerking van spraaksnelheid afhankelijk van de taak is gevonden voor de perceptie van woordgrenzen (hoofdstuk 5) en voor de perceptie van klemtoon (oogbewegingstaak in hoofdstuk 3 en classificatietaak in hoofdstuk 4). De resultaten voor klemtoon en woordgrenzen in de oogbewegingsexperimenten zijn echter niet gelijk. De woordgrenzen in hoofdstuk 5 worden afhankelijk van de spraaksnelheid van de voorafgaande context waargenomen, maar geen dergelijke effecten worden voor de perceptie van klemtoon gevonden (hoofdstuk 3). Een mogelijke verklaring voor dit verschil is het verschil in de temporele eigenschappen van klemtoonpatronen en woordgrenzen. De verwerking van klemtoon eist mogelijk een focus op de volgorde van beklemtoonde en niet-beklemtoonde lettergrepen in het woord. Deze focus op het woord zal dan mogelijke effecten van de voorafgaande context verslulieren. Woordgrenzen, aan de andere kant, zijn niet over de lettergrepen van een woord verdeeld en kunnen dus wel afhankelijk van de spraaksnelheid van de voorafgaande context worden geïnterpreteerd. Er zijn echter tenminste twee resultaten gevonden die deze verklaring tegenspreken. Ten eerste toonden de oogbewegingsexperimenten in hoofdstuk 2 en 3 dat klemtooninformatie meteen wordt gebruikt als deze beschikbaar komt. Het is dus niet noodzakelijk om het hele klemtoonpatroon op een woord te horen om klemtoon voor de woordherkenning te gebruiken. Ten tweede toonde de classificatietaak in hoofdstuk 4 dat klemtoonpatronen worden geïnterpreteerd aan de hand van de spraaksnelheid in de voorafgaande context en dat deze invloed van spraaksnelheid gradueel is (aanwijzingen voor klemtoon in relatie met spraaksnelheid waren belangrijker op initiële dan niet-initiële lettergrepen). Een waarschijnlijkere verklaring voor het verschil tussen online effecten van spraaksnelheid op de locatie van klemtoon en woordgrenzen is het verschil in de stimuli. Zoals eerder besproken, was in ieder geval duurinformatie als aanwijzing voor klemtoon aanwezig op de doelwoorden in hoofdstuk 3. De duur van de grensklanken in hoofdstuk 5 worden naar een ambigue duur gezet. De gemeten effecten moesten dus door de interpretatie van de klanken relatief met de spraaksnelheid van de voorafgaande context zijn veroorzaakt. Er volgt een duidelijke voorspelling: als in de stimuli van hoofdstuk 3 de duur van de eerste twee lettergrepen in de woorden naar een ambigue waarde waren aangepast, zullen er effecten van spraaksnelheid op de klemtoonperceptie ook in de oogbewegingsexperimenten worden gevonden.

Deze observaties benadrukken een belangrijk verschil tussen oogbewegingsexperimenten en classificatietaken. In oogbewegingsexperimenten lijkt het nodig te zijn de kritische duur op een ambigue waarde te zetten om online spraaksnelheidseffecten te kunnen meten. In de classificatietaken in hoofdstuk 4 (klemtoon)

en 5 (woordgrenzen) worden ook niet-ambigue klanken door spraaksnelheid beïnvloed. Om een exact beeld te krijgen hoe verschillende akoestische aanwijzingen tijdens het woordherkenningsproces worden verwerkt is het dus noodzakelijk om verschillende combinaties van taken te gebruiken die verschillende fases van het woordherkenningsproces meten. Een volledig begrip van het woordherkenningsproces is afhankelijk van het vergelijken van verschillende stimuli en taken.

Conclusies

In dit proefschrift wordt onderzocht hoe luisteraars temporele informatie in het voortdurend uitbreidende spraaksignaal verwerken. Eerdere literatuur stelde voor dat luisteraars beschikbare akoestische informatie gradueel gebruiken om woorden optimaal te herkennen (Dahan, Magnuson, Tanenhaus, & Hogan, 2000; Norris & McQueen, 2008; Tanenhaus et al., 1995). Het huidige onderzoek toonde aan dat luisteraars ook suprasegmentele informatie over klemtoonlocatie, de locatie van woordgrenzen en informatie over spraaksnelheid op een gradueel optimale manier kunnen gebruiken. Suprasegmentele aanwijzingen voor klemtoonlocatie worden meteen gebruikt om lexicale competitie te moduleren. Als klemtooninformatie eerder beschikbaar is dan segmentele informatie kunnen woorden door klemtooninformatie alleen worden herkend. Dit wijst erop dat luisteraars hetzelfde perceptiemechanisme gebruiken om segmentele en suprasegmentele informatie op een prelexicaal perceptieniveau te verwerken (zie ook Soto-Faraco, Sebastián-Gallés, & Cutler, 2001). Bovendien wordt duurinformatie over klemtoon en woordgrenzen afhankelijk van de spraaksnelheid van de context geïnterpreteerd. Terwijl meer en meer spraaksignaal binnenkomt, vernieuwen luisteraars niet alleen hun hypotheses over de akoestische informatie om lexicale competitie te moduleren, maar ook de inschattingen van de huidige spraaksnelheid. Spraaksnelheidsinformatie wordt dan tijdens de prelexicale fase van het verwerkingsproces gebruikt voor de evaluatie van nieuw beschikbare segmentele en suprasegmentele informatie. Bovendien veronderstelt het vergelijken van nabije en verafgelegen contexten in online en offline taken dat de hoeveelheid context die kan worden gebruikt afhankelijk is van de hoeveelheid verwerkingstijd die luisteraars beschikbaar hebben. Tijdens het woordherkenningsproces besteden luisteraars meer aandacht aan de context die onmiddellijk vooraf gaat aan het doelwoord dat op dat moment wordt verwerkt. Daarom is het noodzakelijk voor een volledig begrip van het woordherkenningsproces niet alleen het gebruik van verschillende akoestische aanwijzingen te onderzoeken, maar ook verschillende taken te vergelijken die

verschillende fases van het woordherkenningsproces onderzoeken. Deze methode wordt in het hier voorgestelde onderzoek gebruikt. Dit proefschrift leverde nieuwe kennis op over de precieze temporele structuur in de verwerking van gesproken woorden.

Curriculum Vitae

Eva Reinisch was born in 1982 in Graz, Austria. She studied general linguistics and Slavic studies at the University of Vienna, Austria. The academic year 2002-2003 she spent at Adam Mickiewicz University in Poznań, Poland, on an ERASMUS scholarship. In 2005 she received her MA (Mag. Phil.) in general linguistics (distinction). In 2006 she was awarded a 3-year scholarship from the Max Planck Society to do her PhD research at the Max Planck Institute for Psycholinguistics in Nijmegen, The Netherlands, where she joined the Language Comprehension Group. Consequently, a grant for the initiation of research collaborations by the German Research Foundation (DFG) allowed her to spend 6 months as visiting scholar at Emory University in Atlanta, GA (USA). Currently she is working as postdoctoral researcher in the Adaptive Listening Group at the Max Planck Institute for Psycholinguistics.

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 O. Schiller
3. Lexical access in the production of ellipsis and pronouns.
   Bernadette M. Schmitt
4. The open-/closed-class distinction in spoken-word recognition.
   Alette Haveman
5. The acquisition of phonetic categories in young infants: A self-organising artificial neural network approach.
   Kay Behnke
6. Gesture and speech production.
   Jan-Peter de Ruiter
7. Comparative intonational phonology: English and German.
   Esther Grabe
8. Finiteness in adult and child German.
   Ingeborg Lasser
9. Language input for word discovery.
   Joost van de Weijer
10. Inherent complement verbs revisited: Towards an understanding of argument structure in Ewe.
    James Essegbey
11. Producing past and plural inflections.
    Dirk Janssen
12. Valence and transitivity in Saliba: An Oceanic language of Papua New Guinea.
    Anna Margetts

13. From speech to words.
Arie van der Lugt

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

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

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

17. Moving eyes and naming objects.
Femke van der Meulen

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

19. Morphology in speech comprehension.
Kerstin Mauth

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

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

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

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

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

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

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

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

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

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

30. A grammar of Jalonke argument structure.
Friederike Lüpke

31. Agrammatic comprehension: An electrophysiological approach.
Marlies Wassenaar

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

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

34. Phonetic and lexical processing in a second language.
Mirjam Broersma

35. Retrieving semantic and syntactic word properties.
Oliver Müller

36. Lexically-guided perceptual learning in speech processing.
Frank Eisner

37. Sensitivity to detailed acoustic information in word recognition.
Keren B. Shatzman

38. The relationship between spoken word production and comprehension.
Rebecca Özdemir

39. Disfluency: Interrupting speech and gesture.
Mandana Seyfeddinipur

40. The acquisition of phonological structure: Distinguishing contrastive from non-contrastive variation.
Christiane Dietrich

41. Cognitive cladistics and the relativity of spatial cognition.
Daniel B.M. Haun

42. The acquisition of auditory categories.
Martijn Goudbeek

43. Affix reduction in spoken Dutch.
Mark Pluymaekers

44. Continuous-speech segmentation at the beginning of language acquisition: Electrophysiological evidence.
Valesca Kooijman

45. Space and iconicity in German Sign Language (DGS).
Pamela Perniss

46. On the production of morphologically complex words with special attention to effects of frequency.
Heidrun Bien

47. Crosslinguistic influence in first and second languages: Convergence in speech and gesture.
Amanda Brown

48. The acquisition of verb compounding in Mandarin Chinese.
Jidong Chen

49. Phoneme inventories and patterns of speech sound perception.
Anita Wagner

50. Lexical processing of morphologically complex words: An information-theoretical perspective.
Victor Kuperman

51. A grammar of Savosavo, a Papuan language of the Solomon Islands.
Claudia Wegener

52. Prososdic structure in speech production and perception.
Claudia Kuzla

53. The acquisition of finiteness by Turkish learners of German and Turkish learners of French: Investigating knowledge of forms and functions in production and comprehension.
Sarah Schimke

54. Studies on intonation and information structure in child and adult German.
Laura de Ruiter

55. Processing the fine temporal structure of spoken words.
Eva Reinisch