euronoise

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

Eva Reinisch, Alexandra Jesse and James McQueen

Max Planck Institute for Psycholinguistics, Postbus 310, 6500AH Nijmegen, Netherlands
eva.reinisch@mpi.nl

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.

1 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 [1]. 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 [2]. Furthermore, Dutch words become lexically distinct earlier if stress is included in the transcription [3].

Dutch listeners take advantage of stress information to resolve lexical competition during word recognition [4]. 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 [5, 6]. Reinisch, Jesse, and McQueen [7] 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 [3]. 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 [1]. 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., [8]). It also modulates the interpretation of duration as a cue to word boundaries (e.g., [9]). 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 [7] 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.

2 Experiment 1

2.1 Method

2.1.1 Participants

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

2.1.2 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 [10] word frequencies.

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 [11]. 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 [12].

2.1.3 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.

2.2 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 [13]. 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.

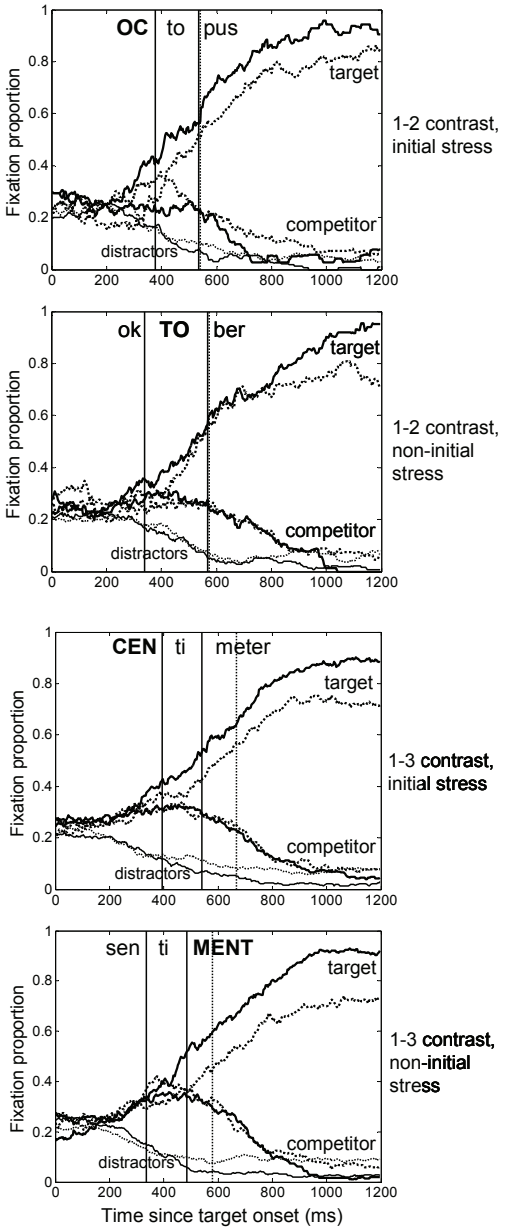

Fig.1 Mean fixation proportions over time to target, competitor, and distractors across conditions. Solid lines represent fast rate, dashed lines slow rate.

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.

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.

2.3 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 [7], 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.

3 Experiment 2

3.1 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 [13] 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.

3.2 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).

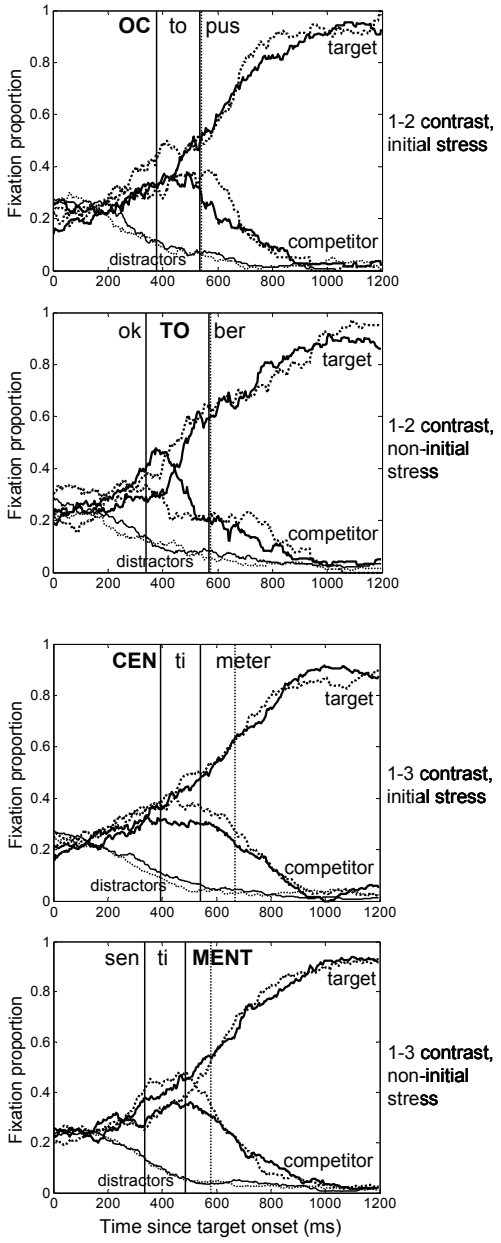

Fig.2 Mean fixation proportions over time to target, competitor, and distractors across conditions**.** Solid lines represent fast rate, dashed lines slow rate.

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.

4 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.

References

[1] A. M. C. Sluijter, V. van Heuven, "Spectral balance as an acoustic correlate of linguistic stress", *J. Acoust. Soc. Am.* 100, 2471-2485 (1996)

[2] A. Cutler, D. Pasveer, "Explaining cross-linguistic differences in effects of lexical stress on spoken-word recognition", in: R. Hoffmann & H. Mixdorff [eds], *Proc. of Speech Prosody*, 237-240 (2006)

[3] V. van Heuven, P. Hagman, "Lexical statistics and spoken word recognition in Dutch", in: P. Coopmans, A. Hulk [eds], *Linguistics in the Netherlands*, 59-68 (1988)

[4] W. van Donselaar, M. Koster, A. Cutler, "Exploring the role of lexical stress in lexical recognition", *Quarterly J. Exp. Psychol. 85A(2)*, 251-273 (2005)

[5] R. M. Cooper, "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", *Cog. Psychol. 6*, 84-107 (1974).

[6] P. D. Allopenna, J. S. Magnuson, M. K. Tanenhaus, "Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models", *J. Mem. Lang., 38*, 419-439 (1998)

[7] E. Reinisch, A. Jesse, J. M. McQueen, "The strength of lexical competition depends on the presence of first-syllable stress", *Proc. of Interspeech*, in press (2008)

[8] Q. Summerfield, "Articulatory rate and perceptual constancy in phonetic perception", *JEP:HPP 7*, 1074-1095 (1981)

[9] B. H. Repp, A. M. Liberman, T. Eccardt, D. Pesetsky, "Perceptual integration of acoustic cues for stop, fricative, and affricate manner", *JEP:HPP 4*, 621-637 (1978)

[10] H. Baayen, R. Piepenbrock, L. Gulikers, "The CELEX Lexical Database (CD-ROM)", Philadelphia, PA: Linguistic Data Consortium (1995)

[11] P. Boersma, D. Weenink, "PRAAT, version 4.6.12", www.praat.org (2007)

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

[13] E. Matin, K. C. Shao, K. R. Boff, "Saccadic overhead: Information-processing time with and without saccades", *Perc. & Psychophysics 53*, 372-380 (1993)