This article was downloaded by: *[Max Planck Institute for Psycholingistic]*
On: *18 March 2010*
Access details: *Access Details: [subscription number 918399536]*
Publisher *Psychology Press*
Informa Ltd Registered in England and Wales Registered Number: 1072954 Registered office: Mortimer House, 37-41 Mortimer Street, London W1T 3JH, UK

The Quarterly Journal of Experimental Psychology
Publication details, including instructions for authors and subscription information:
http://www.informaworld.com/smpp/title~content=t716100704

Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately

Eva Reinisch [a]; Alexandra Jesse [a]; James M. McQueen [ab]
[a] Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands [b] Radboud University Nijmegen, Nijmegen, The Netherlands

First published on: 17 August 2009

To cite this Article Reinisch, Eva, Jesse, Alexandra and McQueen, James M.(2010) 'Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately', The Quarterly Journal of Experimental Psychology, 63: 4, 772 — 783, First published on: 17 August 2009 (iFirst)

To link to this Article: DOI: 10.1080/17470210903104412

URL: http://dx.doi.org/10.1080/17470210903104412

PLEASE SCROLL DOWN FOR ARTICLE

Full terms and conditions of use: http://www.informaworld.com/terms-and-conditions-of-access.pdf

This article may be used for research, teaching and private study purposes. Any substantial or systematic reproduction, re-distribution, re-selling, loan or sub-licensing, systematic supply or distribution in any form to anyone is expressly forbidden.

The publisher does not give any warranty express or implied or make any representation that the contents will be complete or accurate or up to date. The accuracy of any instructions, formulae and drug doses should be independently verified with primary sources. The publisher shall not be liable for any loss, actions, claims, proceedings, demand or costs or damages whatsoever or howsoever caused arising directly or indirectly in connection with or arising out of the use of this material.

THE QUARTERLY JOURNAL OF EXPERIMENTAL PSYCHOLOGY
2010, 63 (4), 772–783

Psychology Press
Taylor & Francis Group

Early use of phonetic information in spoken word recognition: Lexical stress drives eye movements immediately

Eva Reinisch and Alexandra Jesse
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

James M. McQueen
Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands, and Radboud University Nijmegen, Nijmegen, The Netherlands

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 noninitially stressed words. Listeners recognize words by immediately using all relevant information in the speech signal.

Keywords: Spoken word recognition; Lexical stress; Eye tracking.

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 the information comes available (Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Norris & McQueen, 2008; Tanenhaus, Spivey-Knowlton, Eberhard, & Sedivy, 1995).

Correspondence should be addressed to Eva Reinisch, Max Planck Institute for Psycholinguistics, P.O. Box 310, 6500 AH Nijmegen, The Netherlands. E-mail: Eva.Reinisch@mpi.nl

This study was conducted as part of the first author's PhD project, funded by the Max Planck Society. It was presented at Interspeech 2008, Brisbane, Australia. The third author is a member of the Behavioural Science Institute and of the Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, both at the Radboud University Nijmegen. We thank Simon Liversedge, Uli Frauenfelder, and an anonymous reviewer for comments on previous versions of this paper, and Holger Mitterer for suggesting the point-by-point analyses.

http://www.psypress.com/qjep

DOI:10.1080/17470210903104412

The speech signal, however, provides not only 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 whether 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 than in 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 of 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 postperceptual 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 postperceptual 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; R. M. 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 noninitially 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 noninitial 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 noninitial 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 noninitially 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

A total of 24 Dutch native speakers with no reported hearing problems and normal or corrected-to-normal vision were paid for taking part.

Stimuli

A total of 24 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. A total of 7 pairs had stress on the first or the second syllable (e.g., "OCtopus", "okTOber"; 1−2 stress contrast); 17 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. A total of 6 pairs from each contrast set could be distinguished segmentally at the end of the second syllable, and the others within the first two phonemes of the third syllable. A total of 8 additional word pairs were chosen with similar criteria to serve as fillers and 6 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 were 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 1,200 ms.

Procedure

Participants were seated approximately 60 cm in front of a 32.5 × 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 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 12 practice trials that consisted of 6 displays, each presented twice.

On every trial participants saw a fixation cross for 500 ms centred on the screen. After a period of 200 ms, four printed words appeared for 2,400 ms. All words were presented in monospaced lower-case Lucida Sans Typewriter font, size 20, centred 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 1,200 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 5th 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 analysed. 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 side length 6.3 cm, centred around the middle of each word.

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 programme 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 on 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.

Critical time window

Analyses of variance by participants (F_1) and by items (F_2) were run with stress contrast (1–2 and 1–3 contrast) and stress location (primary stress

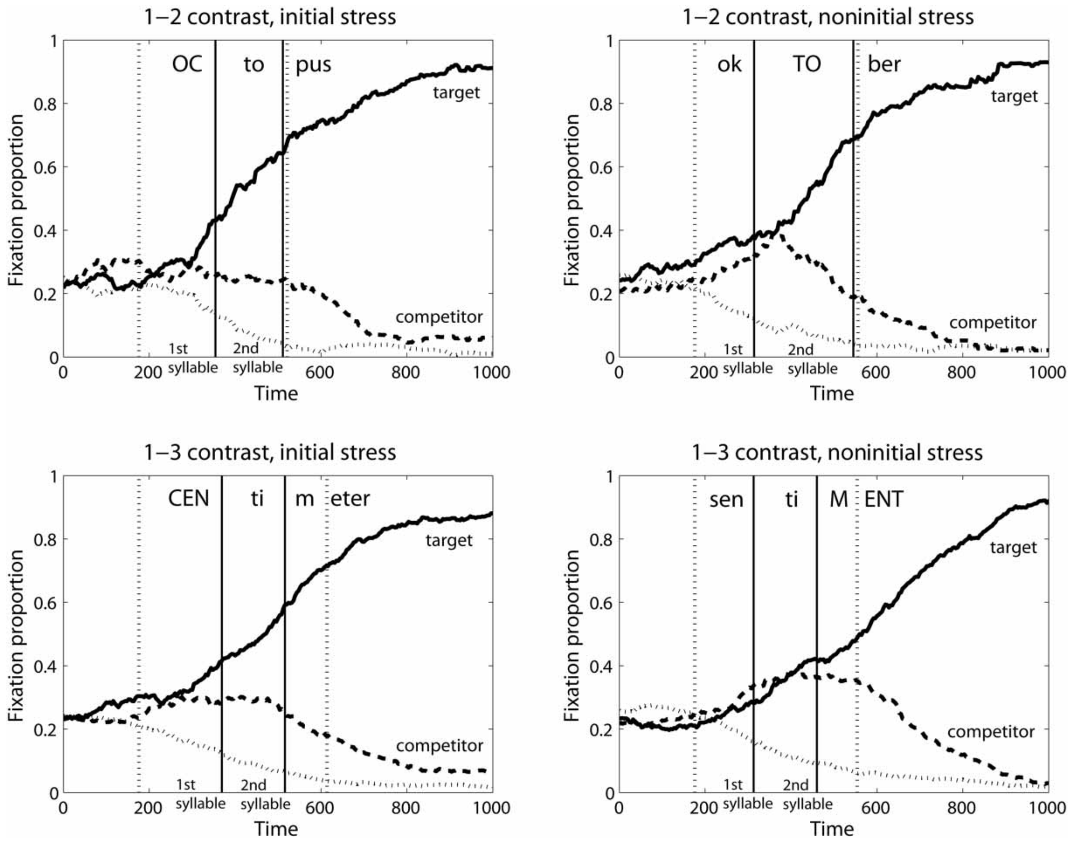

Figure 1. *Fixation proportions over time to target (solid line), competitor (dashed line), and averaged distractors (dotted line) from acoustic target onset. Vertical dotted lines indicate the critical time window (see text for details). Solid vertical lines show normalized and time-shifted syllable 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).*

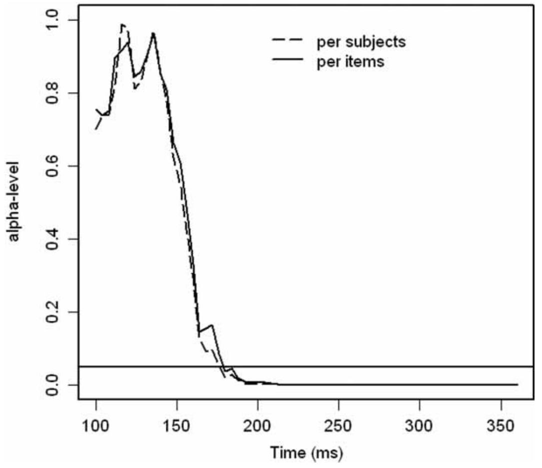

Figure 2. *Point-to-point analysis in consecutive time windows of 4 ms. Plot of p-values (i.e., probability that there is no difference in fixation proportions between distractors and targets or competitors) over time from target onset. The horizontal line indicates the significance level of* $\alpha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

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 noninitially stressed targets from the 1–2 stress contrast received more fixations than noninitially stressed targets from the 1–3 contrast, $F_1(1, 23) = 11.08$, $p < .005$; $F_2(1, 22) = 7.71$, $p < .05$, but there was no difference between the two types of targets with initial stress (F_1 and $F_2 < 1$). 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 noninitial 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 noninitially stressed targets, $F_1(1, 23) = 7.42$, $p < .05$; $F_2(1, 22) = 10.54$, $p < .005$, but not to initially stressed targets (F_1 and $F_2 < 1$). On the second syllable an

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*

			Contrast		*Location*		*Contrast × Location*	
			F	*p*	*F*	*p*	*F*	*p*
Target fixations	Critical time window	$F_1(23)$	0.84	.36	0.003	.96	8.54	<.01
		$F_2(44)$	0.68	.41	0.008	.93	5.19	<.05
	First syllable	$F_1(23)$	1.53	.23	0.003	.96	11.76	<.005
		$F_2(44)$	2.11	.15	0.006	.94	6.59	<.05
	Second syllable	$F_1(23)$	5.69	<.05	0.34	.57	2.44	.13
		$F_2(44)$	6.55	<.05	0.35	.56	2.17	.15
Fixations on competitor minus distractors	Critical time window	$F_1(23)$	0.50	.49	5.91	<.05	1.73	.20
		$F_2(44)$	0.25	.62	3.40	.072	1.23	.27
	First syllable	$F_1(23)$	0.005	.95	0.52	.48	2.20	.15
		$F_2(44)$	0.001	.98	0.37	.55	1.09	.30
	Second syllable	$F_1(23)$	0.29	.6	6.98	<.05	0.91	.35
		$F_2(44)$	0.13	.72	3.80	.058	0.87	.36

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 noninitial stress) was found on the second syllable only. On the first syllable no effect of contrast or location was apparent.

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. 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 root mean square (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 analysed separately but rather pooled with the word pairs from the 1–3 contrast. Paired *t* tests showed that words with initial and noninitial stress differed significantly on all four acoustic measures (see Table 4). As expected, vowels of

Table 2. *Preference of fixations on the target over the competitor per condition in the critical time window*

	1–2 Stress contrast				*1–3 Stress contrast*			
	Initial stress		*Noninitial stress*		*Initial stress*		*Noninitial stress*	
	$t_1(23)$	$t_2(6)$	$t_1(23)$	$t_2(6)$	$t_1(23)$	$t_2(16)$	$t_1(23)$	$t_2(16)$
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

Note: The critical time window was from 180 ms after target onset to the segmental target–competitor divergence point plus 180 ms.

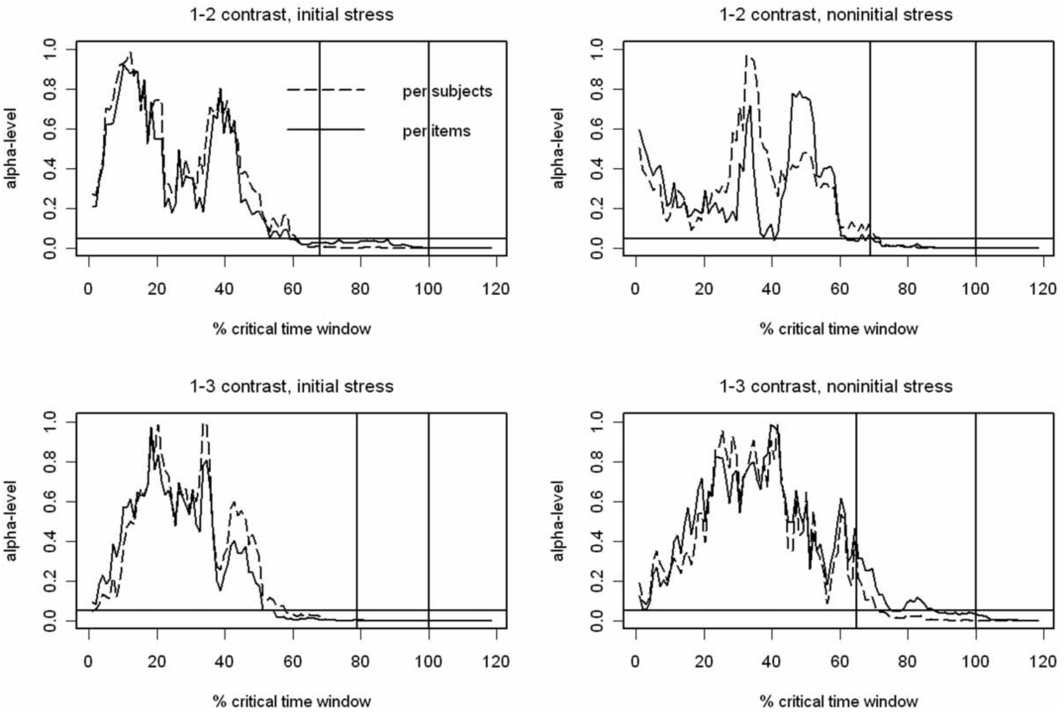

Figure 3. *Point-to-point analysis of target–competitor divergence per condition. Plot of p-values over time in percentage of the critical time window (end of critical time window = 100%). The horizontal line indicates the significance level of α = .05, the left vertical line indicates the average onset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

Table 3. *Time of the onset of the last shared segment of target and competitor and the time at which fixations on the target became more frequent than fixations on the competitor and remained so for at least 20 time slices*

		Onset of last shared segment	*Target–competitor disambiguation in point-to-point analyses*	
Stress contrast	*Stress*	*Time*	t *test*	*Time*
1–2	Initial	68	$t_1(23)$	59
			$t_2(6)$	58
	Noninitial	69	$t_1(23)$	71
			$t_2(6)$	68
1–3	Initial	79	$t_1(23)$	57
			$t_2(16)$	54
	Noninitial	65	$t_1(23)$	70
			$t_2(16)$	86

Note: Time is given as a percentage of the critical time window per condition.
$p < .05$, by participants, t_1, and by items, t_2.

stressed initial syllables were longer, louder, and 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 behavioural measures were calculated. Correlations were performed on difference measures between initially and noninitially stressed words for the acoustic measures and for the eye-movement data from the critical time window. Measures of noninitially 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

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

	Vowel				
Measurement	*Stressed*	*Unstressed*	*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: Spectral tilt is a ratio and therefore does not have a unit.

between items of a pair led to a larger difference in the amount of fixations on the respective targets. No other correlations with target fixation behaviour 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 noninitially 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 behaviour 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 judgements (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 to 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. After a period of 180 ms after target onset, fixations on the segmentally mismatching distractors became less frequent than fixations on target or competitor. After this time, there were three primary results. First, information

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

	Target fixations		*Fixations on competitor minus distractor*		*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: Spectral tilt is a ratio and therefore does not have a unit.

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 noninitially 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 noninitially 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 with regard to stress in that it did not differentially support initially or noninitially stressed words. Words with ambiguous stress cues might generally be recognized more slowly and suffer from more competition than words with nonambiguous stress cues. This was also suggested by Reinisch, Jesse, and McQueen (2008), 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. Noninitially stressed targets from the 1–2 contrast, however, received more fixations than noninitially 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; N. Cooper, Cutler, & Wales, 2002), and 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.

Original manuscript received 22 January 2009
Accepted revision received 29 April 2009
First published online 18 August 2009

REFERENCES

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

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

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.

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

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., 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, Germany: 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.

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.

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

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

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

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.

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

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, The Netherlands: 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.

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			
Contrast	*Initial*	*Frequency*	*Noninitial*	*Frequency*
1–2	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	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 (trainee 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

(*Continued*)