

This article was downloaded by:[Max Planck Inst & Research Groups Consortium]
On: 9 November 2007
Access Details: [subscription number 771335669]
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>

Tracking recognition of spoken words by tracking looks to printed words

James M. Mcqueen^a; Malte C. Viebahn^b

^a Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

^b Westfälische-Wilhelms Universität Münster, Münster, Germany

First Published on: 21 February 2007

To cite this Article: Mcqueen, James M. and Viebahn, Malte C. (2007) 'Tracking recognition of spoken words by tracking looks to printed words', The Quarterly Journal of Experimental Psychology, 60:5, 661 - 671

To link to this article: DOI: 10.1080/17470210601183890

URL: <http://dx.doi.org/10.1080/17470210601183890>

PLEASE SCROLL DOWN FOR ARTICLE

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

This article maybe 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.

Short article

Tracking recognition of spoken words by tracking looks to printed words

James M. McQueen

Max Planck Institute for Psycholinguistics, Nijmegen, The Netherlands

Malte C. Viebahn

Westfälische-Wilhelms Universität Münster, Münster, Germany

Eye movements of Dutch participants were tracked as they looked at arrays of four words on a computer screen and followed spoken instructions (e.g., “Klik op het woord buffel”: Click on the word buffalo). The arrays included the target (e.g., *buffel*), a phonological competitor (e.g., *buffer*, *buffer*), and two unrelated distractors. Targets were monosyllabic or bisyllabic, and competitors mismatched targets only on either their onset or offset phoneme and only by one distinctive feature. Participants looked at competitors more than at distractors, but this effect was much stronger for offset-mismatch than onset-mismatch competitors. Fixations to competitors started to decrease as soon as phonetic evidence disfavouring those competitors could influence behaviour. These results confirm that listeners continuously update their interpretation of words as the evidence in the speech signal unfolds and hence establish the viability of the methodology of using eye movements to arrays of printed words to track spoken-word recognition.

Spoken-word recognition is fast and efficient. Listeners usually have little trouble recognizing the words that speakers intend and do so seemingly instantaneously. This makes it hard to study the underlying processes. Researchers have therefore had to be resourceful in developing suitable methodologies (Grosjean & Frauenfelder, 1996). One successful technique for examining the temporal dynamics of the word recognition process is eye tracking: Eye movements to arrays

of pictures on a computer screen are recorded as participants listen to continuous speech (e.g., Allopenna, Magnuson, & Tanenhaus, 1998). A problem with this method, however, is that critical stimuli have to be picturable. For example, it is difficult to explore the recognition of words that cannot be portrayed in simple visual displays, such as truth and beauty. This limitation imposes strong constraints on material selection. In the present study, therefore, we explored an alternative

Correspondence should be addressed to James M. McQueen, Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands. E-mail: james.mcqueen@mpi.nl

This work was conducted while the second author was a visiting research student at the Max Planck Institute for Psycholinguistics, Nijmegen. It was presented at the Tagung experimentell arbeitender Psychologen (TeaP), Mainz, March 2006. We thank three reviewers for their constructive advice.

methodology that does not have this limitation: tracking looks to arrays of printed words.

We examined the influence of mismatching phonetic evidence on spoken-word recognition. Explorations of what happens when listeners hear mispronunciations using a variety of techniques, including cross-modal priming (e.g., Marslen-Wilson & Zwitserlood, 1989; Soto-Faraco, Sebastián-Gallés, & Cutler, 2001), phoneme monitoring (e.g., Connine, Titone, Deelman, & Blasko, 1997; Frauenfelder, Scholten, & Content, 2001), and eye tracking with pictures (e.g., Allopenna et al., 1998), have shown that negative evidence counts more heavily against a word than positive evidence counts in its favour. A mispronunciation involving just one phoneme can therefore interfere substantially with word recognition. But the effects of mismatch depend on at least four factors. First, word-initial mismatch interferes with lexical access more than word-final mismatch (Allopenna et al., 1998). Second, a mispronunciation disrupts the recognition of short words more than that of long words (compare Connine, Blasko, & Titone, 1993, with Gow, 2001). Third, mismatch effects depend on lexical similarity: A mispronunciation that creates another word has more severe consequences than one that creates a nonword (Marslen-Wilson, Moss, & van Halen, 1996). Fourth, the greater the phonetic similarity of the mispronounced and correct segments, the smaller the disruption is (Connine et al., 1993, 1997; Marslen-Wilson et al., 1996). This experimental evidence, however, is still rather fragmentary. For example, effects of mismatch in long and short words have not been directly compared. Phonemic mismatch therefore provided a suitable test of the viability of printed-word eye tracking. There were sufficient previous findings for clear predictions, but there was also the need to compare effects within a single experiment.

We present a printed-word variant of the Allopenna et al. (1998) study. In that study, American English participants' eye movements were recorded as they looked at a computer screen containing line-drawings of four objects.

The participants' task was to follow spoken instructions to move one of the objects with the computer mouse (e.g., "Pick up the beaker"). The display contained the referent, objects with names beginning or ending in the same way as that of the referent, and an object with a phonologically unrelated name (e.g., a beaker, a beetle, a speaker, and a carriage). Participants looked at the pictures of both types of competitor more than at the unrelated distractors and more at the offset-mismatch competitors (e.g., the beetle) than at the onset-mismatch competitors (e.g., the speaker). Allopenna et al. argued that this pattern of eye fixations reflected the ongoing spoken-word recognition process: As the support for different lexical hypotheses changes over time, the probability of fixations to pictures corresponding to those hypotheses also changes. Offset-mismatch competitors are thus strong candidates until the mismatch has been processed, but onset-mismatch competitors are still considered in spite of their initial poor fit.

Subsequent eye-tracking experiments using picture displays have confirmed that this methodology is suitable for tracking how phonetic information influences word recognition processes over time (e.g., Dahan, Magnuson, Tanenhaus, & Hogan, 2001; Salverda, Dahan, & McQueen, 2003; Shatzman & McQueen, 2006). The question we asked here was whether these processes could be tracked using displays of printed words. We tracked Dutch participants' eye movements while they looked at an array of four printed words (e.g., the target *buffel*, buffalo, the competitor *buffer*, buffer, and two unrelated words) and followed spoken instructions (e.g., "Klik op het woord *buffel*": Click on the word *buffalo*). As in the Allopenna et al. (1998) study, we compared the effects of offset mismatch (e.g., *buffel*/*buffer*) and onset mismatch (e.g., *rotje*/*lotje*; fire-cracker/lottery ticket). We could thus predict that if eye movements to printed words reflect the same spoken-word recognition processes as those revealed in studies with picture displays, participants should look longer at the offset-mismatch competitors than at the onset-mismatch competitors.

We also made two advances on the Allopenna et al. (1998) study. First, we manipulated word length. Targets and their competitors were either bisyllabic (as in the earlier study) or monosyllabic. As already noted, comparisons across previous studies suggest that a mispronunciation should disrupt word recognition less in a longer word. This is because a single-phoneme mispronunciation is proportionally a smaller mismatch in a longer than in a shorter word. We therefore predicted more looks to bisyllabic than to monosyllabic competitors.

The second advance was that degree of mismatch was controlled. In the Allopenna et al. (1998) study there was no control for phonetic similarity between the onset- and offset-mismatch conditions (e.g., the [sp] of *speaker* does not differ from the [b] of *beaker* to the same degree as the [təɪ] of *beetle* does). In the present study, however, the mismatches always involved only one distinctive feature. Phonemes can be described in terms of distinctive features (Jakobson, Fant, & Halle, 1952) such as those coding voicing and place of articulation. Our use of differences of only one feature (e.g., the /l/ and /r/ in *buffel* and *buffer* differ only in manner of articulation) meant we could examine the effects of the smallest possible difference that still involves a phonemic change. Although the effects of yet smaller changes on word recognition have been reported in other eye-tracking studies with pictures, such studies involved artificial manipulation of the speech signal using techniques such as cross-splicing (Dahan et al., 2001; Salverda et al., 2003; Shatzman & McQueen, 2006). Our manipulation did not require speech editing. We were therefore able to examine the effects of both mismatch position and word length with degree of mismatch controlled to be the smallest possible natural phonemic mispronunciation. It was impossible to design a set of picturable materials in Dutch with this degree of control.

We also took the opportunity to reexamine the time-course of spoken-word recognition. Findings such as those of Allopenna et al. (1998) and those from a wide variety of other studies (reviewed in McQueen, in press) converge on the view that

word recognition is continuous and incremental. That is, as acoustic information unfolds over time it is used rapidly to constrain lexical interpretation. One relatively extreme version of this view is that instantiated in the cohort model (Marslen-Wilson, 1987): As soon as a word mismatches the input, even by a single phoneme, it is thrown out of the cohort of current lexical hypotheses. Allopenna et al.'s results (and those of, e.g., Connine et al., 1997, and Frauenfelder et al., 2001) contradict the strict cohort view, which incorrectly predicts that onset-mismatch words (e.g., *speaker* given the input *beaker*) should never be considered for recognition. These results suggest that while word recognition is still very rapid, it is more graded and continuous with respect to mismatching information than in the original cohort model. Our examination of eye movements to onset-mismatch competitors was thus a further test that spoken-word recognition is graded, as assumed in TRACE (McClelland & Elman, 1986) and Shortlist (Norris, 1994).

Even if mismatching information is not used to constrain lexical search in an all-or-none manner, it should nonetheless be used as rapidly as possible. We tested this by estimating the location in time of the peak of the competitor fixation functions in the offset-mismatch conditions. Note that there may be no clear peak in the onset-mismatch conditions because of the early arrival of the mismatching information. But there ought to be a clear peak in fixations to offset-mismatch competitors because they are plausible interpretations of the input until the arrival of the mismatching phoneme. If word recognition is rapid, looks to offset-mismatch competitors (e.g., *buffer*) should peak (i.e., start to decrease) as the final phoneme of the target (e.g., the /l/ of *buffel*) is being heard.

Method

Participants

A total of 16 Dutch native speakers were paid for participating.

Stimuli

Visual displays on experimental trials consisted of four printed words: a target, a competitor, and two distractors. For each of four between-item conditions, 12 pairs of words were selected from the CELEX database (Baayen, Piepenbrock, & Gulikers, 1995). Each word pair consisted of a target and a competitor differing only in one phoneme. A total of 24 pairs were monosyllabic, and 24 were bisyllabic (see Appendix). Within each length condition half of the target–competitor pairs differed only in their onset phoneme, and half differed only in their offset phoneme. These mismatching phonemes differed only in one distinctive feature. These pairs of phonemes were the same in 10 of the 12 trials in each condition: five /l/–/r/ (manner), two /n/–/l/ (manner), one /t/–/k/ (place), one /p/–/k/ (place), and one /x/–/s/ (place). The remaining pairs differed across conditions but all involved stop–fricative manner changes. We avoided voicing differences since they may be less informative than manner and place differences in Dutch (Ernestus & Mak, 2004).

Two additional words for each target–competitor pair were assigned to be distractors. The distractors were phonologically dissimilar to each other and to the target–competitor pair. All four words in each set had the same number of letters and phonemes. The monosyllabic sets had 3–5 letters and 3 or 4 phonemes; the bisyllabic sets had 5–8 letters and 5–7 phonemes. Sets were also matched on CELEX estimates of frequency of occurrence. It was not possible to match target, competitor, and distractor frequencies perfectly (see Appendix). Competitors were nevertheless controlled to be consistently lower in frequency than both targets and distractors. Differences in performance across conditions could therefore not be due to any residual frequency differences. In any case, frequency differences among targets, competitors, and distractors were not significant in any of the conditions (all F s < 2.14; all p s > .1). There were no semantic or morphological relationships among the words within each set (other than that some shared affixes).

An additional 48 quadruplets were selected for filler trials. Each set again consisted of two phonologically similar and two unrelated words, but one of the dissimilar words was the target, and the two similar words and the other dissimilar word were distractors. There were again 24 monosyllabic and 24 bisyllabic pairs, and half of each differed only on either their initial or their final phoneme. The words within each filler set were matched in number of phonemes and letters. A further 12 quadruplets, constructed like the experimental and filler stimuli, were selected for practice trials.

Procedure

For each of the target words an instruction sentence was recorded consisting of the phrase “Klik op het woord” (click on the word) plus the target. All sentences were read aloud in random order by a female native speaker of Dutch in a sound-damped booth.

Participants were seated at a comfortable distance from a computer screen. One centimetre on the screen corresponded approximately to 1° of visual arc. Participants were told that they would be fitted with an eye-tracking device and that they should follow instructions to find target words on the screen and click on them with the computer’s mouse. An SMI Eyelink eye-tracking system was then fitted and calibrated. Spoken instructions were presented to the participants through headphones.

The structure of a trial was as follows (see Figure 1). First, a blank screen appeared for 100 ms, followed by a central fixation cross. The instructions began 2.5 s later. The four words appeared on the screen 200 ms before the acoustic onset of the target word. Note that there was a variable pause between the carrier phrase and the target, but this pause was always larger than 200 ms. Participants clicked with the mouse on the target word and thereby initiated the next trial. Participants were put under no time pressure.

The printed words were presented in 24-point Times New Roman font. There were four fixed locations for the words in a two-by-two array. The horizontal distance between the centres of

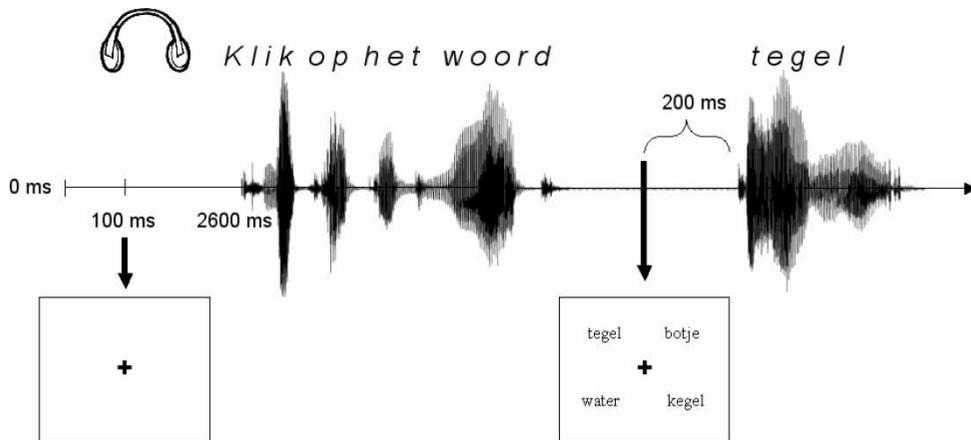


Figure 1. *Experimental procedure.*

the words on the left and right was 9.4 cm. The centres of the upper words were 9.3 cm above those of the lower words. The positions of the four words in each trial were randomized across these four locations.

Four lists were created consisting of one randomization of the 12 practice trials followed by one of four different randomizations of the experimental and filler items. A total of 4 participants were randomly assigned to each list.

Results

The number of fixations to targets, competitors, and distractors was counted in 10-ms bins starting from the acoustic onset of the target and ending 1 s later. A fixation was counted as being directed towards a given word if it fell within a predefined 4.5×4.5 -cm square centred on the middle of the word. All fixations outside these areas were counted as one category. Proportions of fixations to targets, competitors, and distractors were then computed for each bin in each of the four conditions. Mean proportions of fixations over time in each condition are shown in Figure 2.

The proportions of fixations to different types of words started to diverge 200–300 ms after acoustic target onset and had largely converged on the target 500 ms later. The very early preference for

Distractor 2 in the onset-mismatch, monosyllabic condition reflects a strong bias in one trial for the high-frequency distractor *kind*, child. After divergence of the fixation functions, targets were fixated most of all, but competitors were fixated more often than distractors. Following Allopenna et al. (1998), we compared fixation proportions to competitors and distractors in 100-ms bins. The results of analyses of variance (ANOVAs) by participants (F_1) and items (F_2) for each bin in the critical 200–800-ms range are shown in Table 1.

The bisyllabic conditions most closely resemble the conditions tested by Allopenna et al. (1998) and replicate their findings. There were more looks to competitors than to distractors starting early and extending over a broad time window for the offset-mismatch bisyllables. But for the onset-mismatch bisyllables there was only a small effect relatively late (in the 600–700 bin; note that the effect in the 200–300 and 300–400 bins is a reverse effect due to an early preference for distractors; see Figure 2). Note also that, with respect to statistical significance, the effect for onset-mismatch bisyllables is very similar to that found by Allopenna et al. (their effect was significant at the $p < .05$ level on a one-tailed t test by participants in only one 100-ms bin).

The results of these analyses for the new monosyllabic conditions suggest that there were

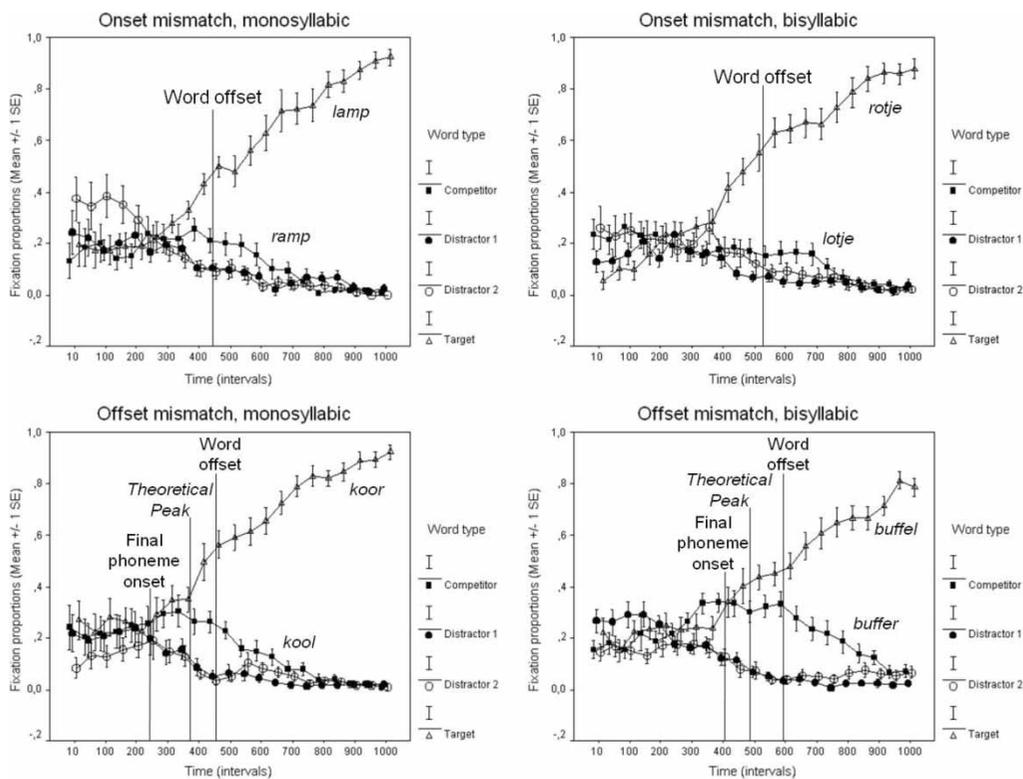


Figure 2. Mean proportion of fixations to printed-word targets, competitors, and distractors, for the 1-s period following acoustic target-word onset. These data, and examples of targets and competitors, are given in each of four conditions: Words were either monosyllabic or bisyllabic, and competitors mismatched with targets in either their onset or offset phoneme. The average acoustic offset of the target word is given in each condition. For each offset-mismatch condition, the average onset of the target word's final phoneme and the location of the theoretical peak of the competitor fixation function are also shown.

fewer looks to competitors than in the bisyllabic conditions and again much stronger competitor effects for offset- than onset-mismatch items. But although direct comparisons of competitor and distractor fixations are required to compare these results with those of the earlier study using pictures, such analyses are questionable since the competitors and distractor fixation proportions are not independent. The primary analysis of the mismatch position and word length manipulations was therefore based on comparisons across conditions of the difference in proportion of fixations to competitors versus distractors. The difference values were computed, either for each participant or for each item, in each condition

over three time windows measured from acoustic target onset: 200–800 ms, 200–500 ms, and 500–800 ms. The 200–800-ms time window is the interval over which fixation proportions to competitors were higher than those to distractors; the smaller windows provide measures of early versus late components of any effects. The tendency to look more at the competitor than at the distractor was stronger when competitors mismatched with targets at offset than when they mismatched at onset: 200–800 ms, $F_1(1, 15) = 12.23$, $p = .003$, $F_2(1, 44) = 10.74$, $p = .002$; 200–500 ms, $F_1(1, 15) = 8.94$, $p = .009$, $F_2(1, 44) = 6.92$, $p = .012$; 500–800 ms, $F_1(1, 15) = 6.80$, $p = .020$, $F_2(1, 44) = 8.12$, $p = .007$. The effect of

Table 1. Analyses of variance by participant and item comparing mean fixation proportions to competitor words with those to the average of the two distractor words in 100-ms bins, from 200 ms to 800 ms after acoustic target word onset, in each of the four conditions

Condition	Test	Time bin ^a					
		200–300	300–400	400–500	500–600	600–700	700–800
Onset mismatch, monosyllabic	$F_1(1, 15)$	1.51, $p = .238$	< 1	1.07, $p = .317$	1.90, $p = .188$	< 1	2.28, $p = .152$
	$F_2(1, 11)$	3.01, $p = .111$	< 1	1.09, $p = .319$	1.24, $p = .290$	< 1	1.01, $p = .336$
Onset mismatch, bisyllabic	$F_1(1, 15)$	2.38, $p = .144$	10.64, $p = .005$	< 1	< 1	4.46, $p = .052$	1.53, $p = .235$
	$F_2(1, 11)$	2.50, $p = .142$	7.63, $p = .019$	< 1	< 1	3.05, $p = .108$	2.31, $p = .156$
Offset mismatch, monosyllabic	$F_1(1, 15)$	< 1	2.95, $p = .107$	13.78, $p = .002$	4.11, $p = .061$	1.77, $p = .204$	< 1
	$F_2(1, 11)$	< 1	3.54, $p = .087$	18.44, $p = .001$	4.02, $p = .070$	1.57, $p = .236$	1.97, $p = .188$
Offset mismatch, bisyllabic	$F_1(1, 15)$	1.86, $p = .193$	4.23, $p = .058$	11.82, $p = .004$	18.44, $p < .001$	10.54, $p = .005$	10.54, $p = .005$
	$F_2(1, 11)$	< 1	2.38, $p = .152$	4.28, $p = .063$	8.04, $p = .016$	15.78, $p = .002$	20.77, $p < .001$

Note: F_1 = participant. F_2 = item.

^aIn ms.

length was weaker. The preference for the competitor over the distractor was stronger for bisyllabic words than for monosyllabic words, but only later in a trial: 200–800 ms, $F_1(1, 15) = 4.31$, $p = .056$, $F_2(1, 44) = 1.36$, $p = .250$; 200–500 ms, $F_1 < 1$, $F_2 < 1$; 500–800 ms, $F_1(1, 15) = 13.60$, $p = .002$, $F_2(1, 44) = 6.19$, $p = .017$. The interaction of the position and length factors was not significant in any analysis.¹

Locations of the peaks in the proportion of fixations to competitors in the offset-mismatch conditions were then estimated. We computed the overall proportion of fixations to the competitors in the two offset-mismatch conditions in

each 10-ms bin in the 200–1,000-ms range and then fitted a logistic power peak function to those data using TableCurve2D (2007). Note that although this function provides a good fit to visual-world data (Scheepers, Keller, & Lapata, in press), we are using it purely descriptively here. We obtained adjusted r^2 values of .99 for the monosyllabic condition and .97 for the bisyllabic condition. One of the four parameters defining this function is its peak location. The estimated peak locations were 376 ms and 483 ms for the monosyllabic and bisyllabic conditions, respectively. As shown in Figure 2, these fell in the middle of the (average) final phoneme in each condition.

¹ A reviewer pointed out that the length effect could reflect the fact that the frequency bias favouring the distractors over the competitors was stronger in the monosyllabic than the bisyllabic conditions (see Appendix). Correlations were therefore performed for each condition between the difference in frequency between the competitor and the mean of the two distractors on the one hand and the difference in proportion of fixations to competitor and the mean of the two distractors on the other. There was a significant correlation only in the onset-mismatch monosyllabic condition, 200–800 ms, $r(11) = .67$, $p = .017$, and limited to the earlier time window: 200–500 ms, $r(11) = .68$, $p = .015$; 500–800 ms, $r(11) = .36$, $p = .258$. Further analysis revealed that this correlation was due to one stimulus set (that with the distractor *kind*). When this item was removed, the correlation was no longer significant but the pattern of results in the ANOVAs was identical (i.e., no effect of length overall or in the earlier time window, but a significant effect in the later time window). The length effect is therefore not due to a frequency effect, nor is an effect in the early time window masked by one.

Discussion

In a variant of the visual-world paradigm in which listeners followed spoken instructions to find target words in arrays of four printed words, listeners fixated the orthographic forms of phonologically related competitors of the targets more often than forms corresponding to phonologically unrelated words. This is consistent with the widely held view that spoken-word recognition entails the simultaneous evaluation of multiple lexical hypotheses (McQueen, in press). Phonological competitors are plausible alternative interpretations of the current speech material and thus attracted visual attention as listeners searched for the targets.

Ultimately, however, listeners found the target words and thus also rejected the competitor interpretations. But the fate of the competitors was different across conditions. There were more looks to offset-mismatch competitors (e.g., *buffer*, given the target *buffel*) than to onset-mismatch competitors (e.g., *lotje*, given the target *rotje*). Our results thus replicate and strengthen the findings of Allopenna et al. (1998), using tighter experimental control. The pattern found by Allopenna et al. could have been because their onset- and offset-mismatch stimuli were not equated in degree of phonetic mismatch (and/or because the stimuli were presented multiple times). Here, mismatches always involved one distinctive feature and were matched across conditions (and stimuli were not repeated). Our results thus confirm that there is an effect of mismatch location. Competitors that begin in the same way as targets and mismatch later are temporarily stronger candidates than those that already differ from targets at their onset.

We also observed an effect of word length. Although the literature indicates that mismatch tends to disrupt the recognition of short words more than long words, this has not previously been shown within an experiment. We found more competitor fixations in the bisyllabic conditions than in the monosyllabic conditions; this effect was statistically significant only in

the 500–800-ms time window. The effects of mismatching material thus appear to be proportional to the amount of matching material. One mispronounced phoneme in a short word is proportionally more negative evidence than in a long word. Monosyllabic competitors were thus more fully overwhelmed by their targets than bisyllabic competitors were. The slight increase in proportion of fixations to bisyllabic competitors with onset mismatch late in time is suggestive of a weak recovery of these competitors, as the negative evidence of the initial phoneme is counteracted by the build-up of positive support from the following matching phonemes in these relatively long words. There is no hint of this kind of recovery for the monosyllabic onset-mismatch competitors, presumably because they have proportionally less positive support.

We argue that listeners rapidly and continuously update their interpretation of words given the evidence in the unfolding speech signal. The smallest possible natural segmental mispronunciation is one involving a change of only one distinctive feature. This minimal change, however, was still enough to have a strong effect on behaviour. Onset mismatch immediately counted heavily against competitors. That is, from as early as there were differential looks, the targets in the onset-mismatch conditions were already the best bets. Hallett (1986) estimates that 200 ms is required for programming a saccade. The increase in looks to targets in the onset-mismatch conditions began less than 400 ms after acoustic onset of the targets and was thus very early. This rapidity of target dominance is consistent with the cohort model view of incremental speech processing (Marslen-Wilson, 1987). However, as in the Allopenna et al. (1998) picture eye-tracking study, there was some evidence, at least for the bisyllabic items, that onset-mismatch competitors were considered for recognition. Though this evidence was weak, it is consistent with other findings (e.g., Connine et al., 1997; Frauenfelder et al., 2001) that suggest that, contrary to the strict cohort model account but as assumed in TRACE (McClelland & Elman, 1986) and

Shortlist (Norris, 1994), onset mismatch is not sufficient to block lexical access completely. The word recognition system appears to tolerate some mismatching phonetic evidence probably so that it can recover from mispronunciations (e.g., if the listener hears *peaker*, the speaker probably intended the word *beaker*).

Lexical interpretation is nevertheless rapid, as confirmed by the offset-mismatch data. Listeners looked most at the offset-mismatch competitors as they were hearing the target's final phoneme. But there was coarticulatory acoustic-phonetic information specifying the final phoneme earlier than the segmentation points labelled in Figure 2 (e.g., formant-transition information in vowels preceding final consonants). The change in eye movement behaviour was thus as early as it could be: Listeners were already starting to look away 200 ms after the arrival of information inconsistent with the competitor. Word recognition is thus keenly time locked to the information in the speech signal, such that, moment by moment, the most plausible lexical interpretation is favoured.

These results establish the viability of the printed-word variant of the visual-world paradigm. They are consistent with findings using displays of pictures (Alloppenna et al., 1998) and with the results from a variety of other tasks. This convergence suggests both that our results reflect spoken-word recognition processes rather than task-specific processes and that the use of printed-word displays is a valid technique. Much remains to be done before we fully understand how phonological and orthographic information is combined to determine search over an array of printed words (see, e.g., Huettig & McQueen, in press). Nevertheless, the printed-word visual-world paradigm has considerable promise. A problem with picture-based eye-tracking experiments is that pictures may have ambiguous labels. It is thus often necessary to carry out norming studies to establish that pictures are named consistently. This problem does not arise with printed-word displays. The major problem in the design of visual-world experiments using pictures, however, is that all

critical stimuli must be picturable. This constraint vanishes with printed-word displays, making it much easier to design controlled sets of materials.

Original manuscript received 9 August 2006

Accepted revision received 6 December 2006

First published online 21 February 2007

REFERENCES

- Alloppenna, 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, R. H., Piepenbrock, R., & Gulikers, L. (1995). *The CELEX Lexical Database* [CD-ROM]. Philadelphia: University of Pennsylvania: Linguistic Data Consortium.
- Connine, C. M., Blasko, D. G., & Titone, D. (1993). Do the beginnings of spoken words have a special status in auditory word recognition? *Journal of Memory and Language*, *32*, 193–210.
- Connine, C. M., Titone, D., Deelman, T., & Blasko, D. (1997). Similarity mapping in spoken word recognition. *Journal of Memory and Language*, *37*, 463–480.
- 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.
- Ernestus, M., & Mak, W. M. (2004). Distinctive phonological features differ in relevance for both spoken and written word recognition. *Brain and Language*, *90*, 378–392.
- Frauenfelder, U., Scholten, M., & Content, A. (2001). Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition. *Language and Cognitive Processes*, *16*, 583–607.
- Grosjean, F., & Frauenfelder, U. (Eds.). (1996). *A guide to spoken word recognition paradigms*. Hove, UK: Psychology Press.
- Gow, D. W. (2001). Assimilation and anticipation in continuous spoken word recognition. *Journal of Memory and Language*, *45*, 133–159.

- Hallett, P. E. (1986). Eye movements. In K. R. Boff, L. Kaufman, & J. P. Thomas (Eds.), *Handbook of perception and human performance* (Ch. 10, pp. 1–112). New York: Wiley.
- Huettig, F., & McQueen, J. M. (in press). The tug of war between phonological, semantic and shape information in language-mediated visual search. *Journal of Memory and Language*.
- Jakobson, R., Fant, C. G. M., & Halle, M. (1952). *Preliminaries to speech analysis: The distinctive features and their correlates*. Cambridge, MA: MIT Press.
- Marslen-Wilson, W. D. (1987). Functional parallelism in spoken word-recognition. *Cognition*, *25*, 71–102.
- Marslen-Wilson, W., Moss, H. E., & van Halen, S. (1996). Perceptual distance and competition in lexical access. *Journal of Experimental Psychology: Human Perception and Performance*, *22*, 1376–1392.
- Marslen-Wilson, W. D., & Zwitserlood, P. (1989). Accessing spoken words: The importance of word onsets. *Journal of Experimental Psychology: Human Perception and Performance*, *15*, 576–585.
- McClelland, J. L., & Elman, J. L. (1986). The TRACE model of speech perception. *Cognitive Psychology*, *10*, 1–86.
- McQueen, J. M. (in press). Eight questions about spoken-word recognition. In M. G. Gaskell (Ed.), *The Oxford handbook of psycholinguistics*. Oxford, UK: Oxford University Press.
- Norris, D. (1994). Shortlist: A connectionist model of continuous speech recognition. *Cognition*, *52*, 189–234.
- 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.
- Scheepers, C., Keller, F., & Lapata, M. (in press). Evidence for serial coercion: A time course analysis using the visual-world paradigm. *Cognitive Psychology*.
- Shatzman, K. B., & McQueen, J. M. (2006). Segment duration as a cue to word boundaries in spoken-word recognition. *Perception & Psychophysics*, *68*, 1–16.
- Soto-Faraco, S., Sebastián-Gallés, N., & Cutler, A. (2001). Segmental and suprasegmental mismatch in lexical access. *Journal of Memory and Language*, *45*, 412–432.
- Table Curve 2D (2007). [computer software]. Retrieved from <http://www.systat.com>. San Jose, CA: Systat Software, Inc.

APPENDIX

Experimental items (with mean frequencies per million words in parentheses)

<i>Target</i>	<i>Competitor</i>	<i>Distractor</i>		<i>Target</i>	<i>Competitor</i>	<i>Distractor</i>	
		<i>1</i>	<i>2</i>			<i>1</i>	<i>2</i>
<i>Onset mismatch, monosyllabic words</i>				<i>Offset mismatch, monosyllabic words</i>			
rat	lat	pop	hek	koor	kool	haas	teek
lamp	ramp	heks	wind	tol	tor	rem	lap
lijst	rijst	kaart	fiets	kier	kiel	doek	jack
land	rand	wolk	kind	baal	baar	toog	reep
lijm	rijm	moer	pier	boer	boel	wier	lijm
nat	lat	bel	zon	pijn	pijl	soep	duim
nek	lek	dak	pet	maal	maan	veer	been
taak	kaak	muur	boom	beek	beet	paard	loon
kaal	paal	geel	vuur	hak	hap	pot	lam
sok	gok	bad	non	boos	boog	heet	vaas
put	fut	bek	tin	hart	hars	berg	worm
kaas	gaas	heer	boon	leeg	leek	vaag	rook
(77)	(15)	(62)	(125)	(85)	(18)	(42)	(32)
<i>Onset mismatch, bisyllabic words</i>				<i>Offset mismatch, bisyllabic words</i>			
rotje	lotje	veter	panel	buffel	buffer	koppel	visser
lente	rente	zetel	tasje	wimper	wimpel	handel	mentor
liedje	rietje	wekker	zuivel	poeder	poedel	fokker	buisje
landen	randen	bergen	gaatje	dader	dadel	koker	vezel
reder	leder	pakje	bende	kiezer	kiezel	palmen	varken
leger	neger	zadel	rover	banaal	banaan	geweer	legaal
leven	neven	maatje	makker	metaal	methaan	piloot	kanaal
tegel	kegel	botje	water	bereid	bereik	retour	geheim
pater	kater	bakje	meter	voorwerp	voorwerk	handdruk	deurknop
sekte	gekke	pasje	notie	gewas	gewag	galop	debat
teder	ceder	wezel	ledig	geblaas	geblaas	vandaal	winnaar
keurig	geurig	woning	koepel	geloof	geloop	moreel	piraat
(58)	(6)	(13)	(46)	(22)	(3)	(15)	(18)