Changing only the probability that spoken words will be distorted changes how they are recognized[a]

James M. McQueen[b]

Donders Institute for Brain, Cognition and Behaviour, Centre for Cognition, and Behavioural Science Institute, Radboud University Nijmegen, Postbus 9104, 6500 HE Nijmegen, The Netherlands

Falk Huettig[c]

Max Planck Institute for Psycholinguistics, Postbus 310, 6500 AH Nijmegen, The Netherlands

An eye-tracking experiment examined contextual flexibility in speech processing in response to distortions in spoken input. Dutch participants heard Dutch sentences containing critical words and saw four-picture displays. The name of one picture either had the same onset phonemes as the critical word or had a different first phoneme and rhymed. Participants fixated on onset-overlap more than rhyme-overlap pictures, but this tendency varied with speech quality. Relative to a baseline with noise-free sentences, participants looked less at onset-overlap and more at rhyme-overlap pictures when phonemes in the sentences (but not in the critical words) were replaced by noises like those heard on a badly tuned AM radio. The position of the noises (word-initial or word-medial) had no effect. Noises elsewhere in the sentences apparently made evidence about the critical word less reliable: Listeners became less confident of having heard the onset-overlap name but also less sure of having *not* heard the rhyme-overlap name. The same acoustic information has different effects on spoken-word recognition as the probability of distortion changes. © *2012 Acoustical Society of America*. [DOI: 10.1121/1.3664087]

PACS number(s): 43.71.Es [PEI]   Pages: 509–517

I. INTRODUCTION

Speech is inherently variable. For example, talkers speak differently (e.g., because of regional-accent variation), the way words sound changes across different sentential positions (e.g., because of effects of intonation), and sentences vary in their acoustic quality (compare, e.g., those read aloud in a high-fidelity audio book with those spoken in the adverse conditions of a conversation in a busy train station). In order to comprehend speech, listeners have to recognize words in spite of all this variability. They should therefore be able to adapt dynamically to the requirements of different listening situations—good and bad. Here, we tested new limits on listeners' ability to make these dynamic adjustments.

We asked how listeners adjust to speech that is partially masked by noise, such as when they hear Neil Armstrong's announcement of the first landing on the moon. The question was whether the same acoustic evidence (e.g., the word "*leap*") can have a different effect on speech comprehension if the probability of distortion of that word changes because of noises occurring in other parts of the input (e.g., in "*one small *tep*"). There have been many studies of the effects of speech distortion, but in these cases the words being recognized have themselves been disrupted [e.g., by a second

stream of speech (Cherry, 1953), by interruptions (Miller and Licklider, 1950), or by background noise or reverberation (Payton *et al.*, 1994)]. Here, we hold constant the words that we measure the recognition of, and manipulate whether disruptions occur elsewhere. Is the listener able to adjust how the same acoustic information is processed as a function only of changes in the *probability* that that information may be more or less reliable?

Because speech unfolds over time, speech recognition is incremental (Marslen-Wilson and Welsh, 1978; Zwitserlood, 1989). As acoustic information arrives, it immediately constrains the search amongst candidate word hypotheses (Reinisch *et al.*, 2010). Candidates that continue to match the available information keep being considered; those which no longer match tend to be discarded. Incrementality was central to the Cohort model (Marslen-Wilson and Welsh, 1978) and remains a core assumption in subsequent models of spoken-word recognition [e.g., TRACE (McClelland and Elman, 1986), Shortlist (Norris, 1994)]. A related concept is that of optimality in speech recognition (Norris and McQueen, 2008). According to Anderson's (1990) Principle of Rationality, cognitive systems should be optimized with respect to their goals. In the case of spoken-word recognition, the perceptual system should, according to this view, always make the best guess it can, given the available information, about what word is currently being spoken.

These notions of incrementality and optimality together predict that information at the beginning of a word should tend to weigh more heavily in the evaluation of lexical candidates than information at the end of a word. Onset information should tend to constrain lexical search more than

offset information both because it arrives earlier in time and because uncertainty about the spoken word is larger when less of it has been heard. Experimental evidence has shown that this is indeed the case. In a visual-world eye-tracking study, for example, Allopenna *et al.* (1998) showed that a competitor word which begins in the same way as a word that is actually spoken (e.g., the competitor *beetle*, when *beaker* is spoken) competes for recognition more strongly, and for longer, than a competitor which rhymes (e.g., *speaker*). As the listener hears the beginning of *beaker*, *beetle* is as likely an interpretation as *beaker* itself, but *speaker*, due to the mismatching [sp], is already an unlikely hypothesis. The later mismatch between *beetle* and *beaker* cannot prevent *beetle* from being considered as a lexical hypothesis earlier on, and the later rhyming match between *speaker* and *beaker* is unlikely ever to undo completely *speaker*'s initial disadvantage. Thus, although onset- and rhyme-overlap competitors are both considered by the recognition system, rhyming competitors can never fight as strongly as onset competitors—at least not if the initial mismatch between a rhyme competitor and the spoken word is used immediately [indeed, according to the original Cohort model (Marslen-Wilson and Welsh, 1978), if initial mismatches carry their full weight, rhyme competitors should not fight at all].

Several other lines of evidence from a variety of tasks also suggest that information at or near the beginnings of words is more important in constraining lexical selection than information later in words. First, there is greater tolerance for mismatching phonetic material in word-final than in word-initial position [compare, e.g., Zwitserlood (1989), with Marslen-Wilson and Zwitserlood (1989)]. Second, word-initial mismatch blocks lexical access unless it is phonetically very small (Connine *et al.*, 1993; Frauenfelder *et al.*, 2001; Marslen-Wilson *et al.*, 1996). Third, mispronunciations are detected less when they occur later in words, and, during shadowing, later mispronunciations are more readily restored to their correct form than earlier ones (Marslen-Wilson and Welsh, 1978).

But what if the information in the speech signal happens to be less reliable? As we have just discussed, when the quality of the input is high, rhyme competitors (e.g., *speaker* given the input *beaker*) are weaker candidates than cohort competitors (e.g., *beetle*). But if the acoustic information in a given listening situation is likely to be poorer, matches and mismatches between the current input and stored phonological knowledge should both be treated with more caution. So *beetle* should be a weaker candidate when it is less certain that *beaker* has been heard (and *beaker* should be a weaker candidate too). Interestingly, however, *speaker* should become a stronger candidate, because when the signal is less reliable it becomes less certain that *speaker* has not been heard.

We tested this prediction here, using a variant of the Allopenna *et al.* (1998) eye-tracking design. We focused on the situation of listening to an AM radio broadcast with intermittent interference. A baseline group of Dutch participants heard spoken Dutch sentences as they viewed displays containing four line drawings. There were two sets of materials. One set was based on pairs of words that began in the

same way; the second set was based on rhyming pairs that only differed in their initial sounds. One member of a given pair appeared in the sentence, and the other appeared as one of the drawings (together with three unrelated drawings). The spoken word was not displayed visually. This "target-absent" design (Huettig and Altmann, 2005) provides a sensitive measure of competitor strength. We could thus measure how often participants looked, for example, at *krokodil*, "crocodile" as they heard a sentence containing *krokus*, "crocus," and how often they fixated *hamer*, "hammer," as they heard *kamer*. We predicted that, with acoustically clear sentences, we would replicate the Allopenna *et al.* results (i.e., that there would be more looks, on average, to onset competitors such as *krokodil* than to rhyme competitors such as *hamer*). Other eye-tracking studies have also suggested that onset competitors are stronger candidates than rhyme competitors [e.g., Magnuson *et al.* (2003), with an artificial lexicon, at least as training on the novel words increases; Desroches *et al.* (2006), with 9-year-old children; and McQueen and Viebahn (2007), with printed-word displays].

Two further groups of participants saw the same displays and heard the same sentences, but the sentences were occasionally disrupted by bursts of noise. These disruptions did not occur on the critical words (e.g., *krokus* and *kamer*). They occurred only elsewhere in the sentences, either always at the beginnings of words (for the second group), or always in the middle of words (for the third group). The absence of noise on the critical words themselves meant that differences across groups in fixation behavior could not be attributed to acoustic differences on those words, and hence that such differences would have to reflect an adjustment to the heightened probability of distortion in the sentences. That is, we could test whether the speech-recognition system has the flexibility to adjust how it evaluates exactly the same acoustic information in different listening contexts. We predicted that listeners in the second group, relative to the baseline group, would tend to look less at the onset competitors (e.g., *krokodil*) and more at the rhyme competitors (e.g., *hamer*). The presence of radio noise on other word onsets should make the onsets of the critical words (e.g., *krokus* and *kamer*) less trustworthy, weakening support for onset competitors (and the critical words themselves) but also making it less certain that rhyme competitors had not been spoken.

The third group, with word-medial distortions, allowed us to test the specificity of these predicted perceptual adjustments. One possibility was that the adjustments could be specific to onset position. In other words, the listeners in the second group could learn not only that there are noises in the sentences, but also that they always occur at word onsets, and hence that it is specifically word-onset information that is less reliable. If so, then a different outcome would be expected if the distortions were always in medial position. Under these conditions word-onset information is just as reliable, so performance should be like that in the baseline group, and not like that in the group with word-onset distortions. Alternatively, however, there could be a more general adjustment to noise in the carrier sentences. The recognition system could learn that acoustic information is in general less reliable through occasional bursts of noise, but not

J. M. McQueen and F. Huettig: Flexibility in spoken-word recognition

where those noises occur. If so, word-medial noises should have the same effect as word-onset noises, and the two noise groups should perform equivalently.

The comparison of the word-onset and word-medial noise conditions thus allowed us to test whether prior or current experience with speech distortions is more important in determining how adjustments to those distortions are made. In normal AM-radio listening conditions, distortions can appear anywhere. If listeners have learned that distortion position in normal listening is unpredictable, they may not respond differentially to the position-specific distortions in the present experiment. Alternatively, however, it was possible that listeners are sophisticated enough to be able to pick up on the experiment-internal contingencies and hence treat the word-initial and word-medial noise conditions differently. The lack of a difference between the two noise conditions, however, was the more likely outcome, both because cumulative prior experience with distortions in unpredictable conditions is likely to outweigh the effects of positional specificity within a single experiment and because it is known that listeners have difficulty locating extraneous noises [e.g., clicks (Fodor and Bever, 1965) and coughs (Warren, 1970)] in speech.

In summary, we measured fixations to pictures with names which either began in the same way or rhymed with critical words that were heard in sentence contexts. We manipulated whether the sentences were distorted by intermittent noise bursts or not, and the position of those bursts. We tested whether these distortions modulate the way the acoustic information in the critical words influences word recognition, even though this information was itself completely identical across conditions. We thus tested whether, in speech recognition, there is flexibility in the perceptual weight assigned to exactly the same acoustic input appearing in different listening situations.

The critical predictions concerned eye movements to the onset and rhyme competitors in the region 200–400 ms after critical-word onset. We do not predict changes in fixations to the distractor words across conditions: Since they have no phonological overlap with the critical spoken words, they should not receive any significant speech-driven fixations in any condition. The critical time window was chosen for the following reasons. Given that it takes approximately 200 ms to program and launch a saccadic eye movement (see, e.g., Matin *et al*., 1993; Saslow, 1967), fixations earlier than 200 ms cannot reflect responses to the acoustic information in the critical words. In previous studies with clear natural speech (Allopenna *et al*., 1998; McQueen and Viebahn, 2007) there were preferential looks in a 200–400 ms time window to onset competitors, but not to rhyme competitors. By focusing on this region, we could thus ask if this pattern changes when noise is added elsewhere. Furthermore, we predict that the effect of the noise manipulation will be short-lived: As the critical word unfolds over time, it will become clear that it is not distorted. The effect of the heightened probability that the word could have been distorted will thus be outweighed by the physical evidence that it is not distorted. In summary, the 200–400 ms time region covers the first moments of spoken-word recognition and thus will reveal

whether the noise manipulation has an effect on listeners' reliance on word-onset information before the evidence that the critical word is itself not distorted starts to have an effect.

II. METHOD

A. Participants

Participants were 55 paid members of the MPI for Psycholinguistics subject panel, 21 in the baseline group, 16 in the word-onset-noise group, and 18 in the word-medial-noise group. All were native speakers of Dutch, none had known hearing problems, and all had normal or corrected-to-normal visual acuity.

B. Materials

There were 75 visual displays, each paired with a spoken sentence containing a critical word. Visual displays in experimental trials consisted of line drawings of a phonological competitor of the critical word (which was not semantically related to the critical word) and three semantically and phonologically unrelated distractors. There were two experimental conditions, each with 25 trials (see Table I). In the onset-overlap condition the critical word (e.g., *krokus*) overlapped at onset (mean overlap = 4.0 phonemes) with the name of the competitor in the display (e.g., the picture of a crocodile, *krokodil*). In the rhyme-overlap condition the critical word (e.g., *kamer*) overlapped at offset (mean overlap = 3.28 phonemes) with the competitor name and differed only in its initial phoneme (e.g., a hammer, *hamer*).

Each critical word was placed in a neutral sentence in which it was not predictable (e.g., for the critical word *krokus*: *Bestaat er ook een ander woord voor krokus, weet jij dat?*, "Is there another word for crocus, do you know?," and for *kamer*: *Midden in de kamer lag een matras waarop iemand lag te slapen*, "In the middle of the room lay a mattress on which someone lay sleeping"). There were 25 filler trials. The sentences in these trials included a critical word (e.g., *broodrooster*, toaster, in *Toen ze naar binnen keek zag ze een broodrooster op de tafel staan*, "When she looked in she saw a toaster standing on the table") whose referent (e.g., a picture of a toaster) was depicted in the display, along with three unrelated distractors.

C. Stimulus preparation

The sentences were recorded in a sound-damped room by a female native speaker of Dutch who was unaware of the experiment's purpose. Digital recordings of each stimulus were selected and measured using PRAAT (Boersma and Weenink, 2005). The sentences were read with a neutral intonation contour such that, in particular, the critical words were not highlighted. Pictures were black-on-white line drawings.

The same sentences and displays were presented to all three participant groups. The only between-group difference concerned whether there were intermittent noises in the sentences. The baseline group heard the sentences with no noise. For the other two groups, individual phonemes were replaced by noise, either always word-initial phonemes for one group or always word-medial phonemes for the other.

Two, three, or four words were selected per sentence for which one phoneme was replaced by radio noise. For the 25 sentences in each condition (onset overlap, rhyme overlap, and fillers), nine had two noise replacements, eight had three noise replacements, and eight had four noise replacements. The noises could be on any word except for the critical word and the two words before and after it. They were spread around the critical word (e.g., two before and two after). Five different radio noises were selected from Internet databases. The noises were 42262_crk365_mobile_beep1.wav, 495_skiptracer_RadioStatik3.wav, and 30335_EHR_radio_-noise_2.wav from http://www.freesound.org; and whitenoise.wav and radio_whitenoise.wav from http://www.burninwave.com/#whitenoise. These five noises were randomly selected for use within and across the sentences (ranging from 37 to 48 times each). The onsets and offsets of each to-be-replaced phoneme were identified at zero crossings using PRAAT. The phoneme was then substituted by a noise fragment matched in duration to the phoneme. The average loudness of every sentence was then analyzed in ADOBE AUDITION. The loudness of the noises used in a given sentence was set to 80% of the average loudness (in dB) of that sentence. Noises were therefore never louder than the rest of the sentence.

D. Procedure

Participants were seated at a comfortable distance from a computer screen. One centimeter on the display corresponded to approximately $1°$ of visual arc. The eye-tracking system was mounted and calibrated. Eye movements were monitored with an Eyelink 2 system. Spoken sentences were presented through headphones. The four pictures appeared on the screen 1 s before the auditory presentation of the sentence was initiated. Positions of the pictures were randomized across four fixed locations.

Participants were not asked to perform an explicit task. They were told that they should listen to the sentences carefully, that they could look at whatever they wanted to, but that they should not take their eyes off the screen (Huettig and Altmann, 2005). Participants' fixations were thus unconstrained and participants were under no time pressure to perform any action. Each participant was presented with all 75 trials. Experimental and filler trials were presented in random order. A central fixation point appeared on the screen after every five trials, allowing for automatic drift correction in the calibration.

III. RESULTS

Figure 1 shows a time-course graph of fixation proportions at 20 ms intervals to onset (upper panel) and rhyme competitors (lower panel) over the course of the average experimental trial. Zero represents the onset of the critical spoken word. Fixation proportions for the three unrelated distractors were averaged.

A. Overall analyses

We first tested whether the overall data replicated previous studies that compared onset and rhyme competitor effects (Allopenna et al., 1998; Desroches et al., 2006; Magnuson et al., 2003; McQueen and Viebahn, 2007). We computed mean fixation proportions for each type of picture (onset competitor, rhyme competitor, or distractor) over a time interval starting 200 ms after critical-word onset (as an estimate of the earliest point at which fixations could reflect a response based on information in the word; Matin et al., 1993; Saslow, 1967) and extending to 800 ms (as an estimate of when onset and rhyme competitor fixations would likely end, based on the findings of Allopenna et al., 1998, and McQueen and Viebahn, 2007). We calculated the ratio between fixations to a competitor and the sum of the particular competitor and distractor fixation proportions, thus deriving a measure of competitor fixation preference based on eye movements both to the competitors and to the distractors. The mean ratio was then compared to 0.5 (Huettig and McQueen, 2007). A ratio greater than 0.5 reveals that, of all the fixations directed to a particular competitor and the unrelated distractors, the competitors attracted more than half of those fixations.

Figure 2 illustrates that participants in all three groups directed more overt attention to the onset competitors than the unrelated distractors in the 200–800 ms time window. Figure 2 also suggests that the rhyme competitors received more attention than the unrelated distractors in the word-onset and word-medial noise groups but not in the baseline group. One-sample two-tailed t tests (by participants and items) showed that, in the baseline condition, onset competitors [mean ratio 0.65, by participant range 0.56–0.77, $t_1(20) = 13.04$, $p < 0.001$; $t_2(24) = 4.27$, $p < 0.001$] but not rhyme competitors [mean ratio 0.52, range 0.35–0.63, $t_1(20) = 1.15$, $p > 0.1$; $t_2(24) = 0.47$, $p > 0.1$] were fixated significantly more than unrelated distractors. In the word-onset-noise condition, onset competitors [mean ratio 0.63, range 0.48–0.75, $t_1(15) = 5.86$, $p < 0.001$; $t_2(24) = 2.97$, $p < 0.01$] and rhyme competitors [mean ratio 0.56, range 0.41–0.71, $t_1(15) = 2.57$, $p = 0.02$; $t_2(24) = 1.67$, $p = 0.11$] were fixated more than unrelated distractors. In the word-medial-noise condition, onset competitors [mean ratio 0.63, range 0.36–0.77, $t_1(17) = 5.38$, $p < 0.001$; $t_2(24) = 2.57$, $p = 0.02$] and rhyme competitors [mean ratio 0.55, range 0.40–0.66, $t_1(17) = 2.87$, $p = 0.01$; $t_2(24) = 0.92$, $p > 0.1$] were fixated more than unrelated distractors.

The present study thus replicates the previously observed pattern of robust onset effects and smaller, later, and more marginal rhyme effects. Note that although Allopenna et al. (1998) report a significant rhyme-competitor effect, this was based on a one-tailed t-test, over participants only, in a within-item design where each item was repeated multiple times. McQueen and Viebahn (2007) report a statistically marginal rhyme-competitor effect (based on a between-item design with no stimulus repetition, like that used here, but with printed-word displays); this effect was limited to a relatively late time window (600–700 ms after target onset). The rhyme-competitor effect reported by Desroches et al. (2006) was numerically smaller than the onset-competitor effect. Magnuson et al. (2003) report that, as training on novel words in their artificial-lexicon study increased, fixation behavior to onset and rhyme

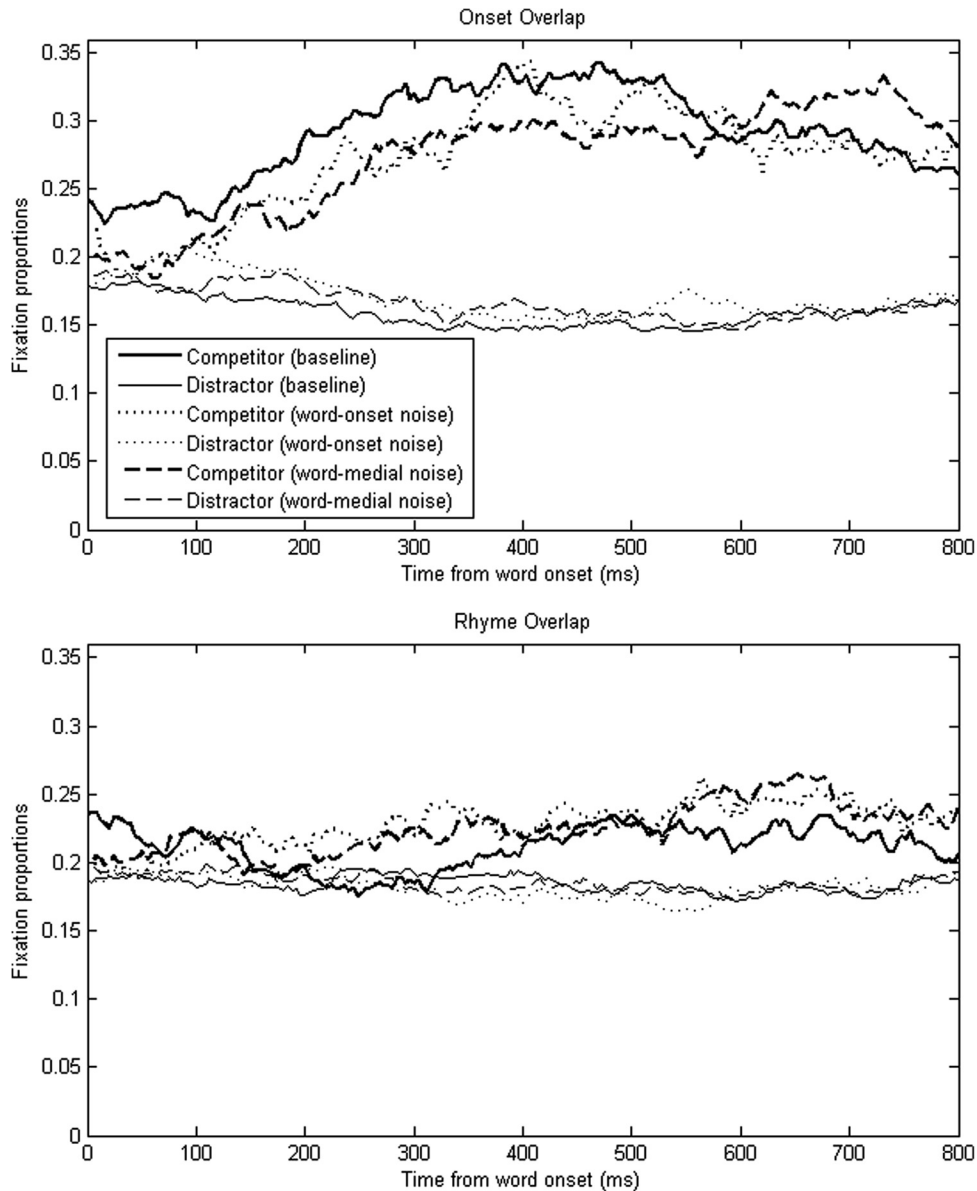J. M. McQueen and F. Huettig: Flexibility in spoken-word recognition

FIG. 1. Mean proportion of fixations to onset competitors (upper panel) and rhyme competitors (lower panel) and their respective distractors in each group (baseline, word-onset noise, and word-medial noise) for the 800 ms period following acoustic onset of the critical word.

competitors more closely resembled that found with real words (i.e., more and earlier looks to onset than to rhyme competitors). It thus appears that, across studies with different designs, including the present one, there are statistically weak but nonetheless replicable effects for rhyme competitors, and much stronger and earlier effects for onset competitors.

The primary purpose of the present study, however, was to investigate the effect on word recognition of intermittent noises occurring elsewhere in the spoken input. Can the presence of such noises weaken the strong onset effect and strengthen the weak rhyme effect? The above-discussed analyses already suggest that this may be the case: The effects for onset competitors became smaller in the noise conditions than in the baseline condition, and the effects for rhyme competitors became larger. This pattern was confirmed in a 2×2 repeated-measures analysis of variance (ANOVA) by participants (F_1) and items (F_2) of fixation ratios in the 200–800 ms region. Noise (no noise in the base-

line condition vs noise in the other two conditions) was a between-subject factor, and competitor type (onset vs rhyme) was a within-subject factor. These analyses revealed a main effect of competitor type [more fixations to onset than to rhyme competitors, $F_1(1,53) = 56.50$, $p < 0.001$; $F_2(1,73) = 11.29$, $p = 0.001$], no effect of noise (F_1 and $F_2 < 1$), and a noise by competitor-type interaction which was significant by participants only [$F_1(1,53) = 4.50$, $p < 0.05$; $F_2(1,73) = 1.10$, $p > 0.1$]. A further 2×2 repeated-measures ANOVA with noise position (word-onset vs word-medial) as between-subject factor revealed no noise position by competitor type interaction [$F_1(1,32) = 0.05$, $p > 0.1$; $F_2(1,48) = 0.19$, $p > 0.1$]. This suggests that the position of the noises did not change the effect that the noises had on word recognition.

Additional analyses that took a different approach to the data confirmed these findings. Linear mixed-effect regression (LMER) analyses compared the competitor fixation proportions in the 200–800 ms region across conditions.
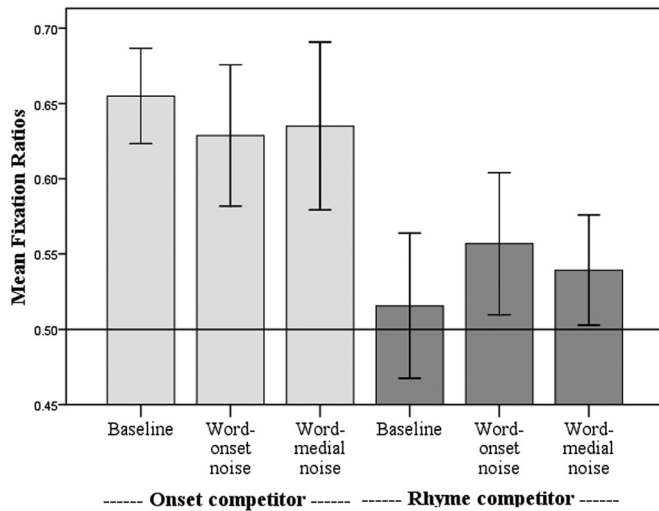
FIG. 2. Mean fixation ratios in the 200–800 ms time window for onset and rhyme competitors in each group (baseline, word-onset noise, and word-medial noise). Error bars represent 95% confidence intervals.
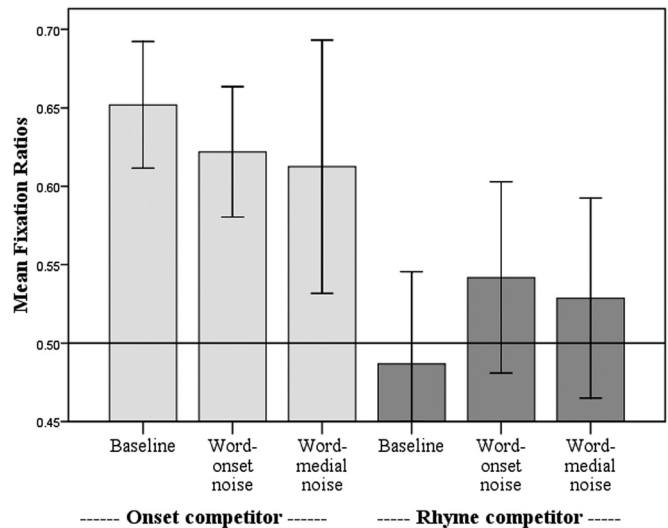

FIG. 3. Mean fixation ratios in the 200–400 ms time window for onset and rhyme competitors in each group (baseline, word-onset noise, and word-medial noise). Error bars represent 95% confidence intervals.

Proportions were transformed to logits prior to the analysis (Barr, 2008). Participants and items were treated as crossed random factors, and p values were estimated using Markov chain Monte Carlo simulations (Baayen *et al.*, 2008). Competitor type was a two-level fixed factor (onset vs rhyme) and noise was a three-level fixed factor (no-noise baseline vs word-onset noise vs word-medial noise). There was a significant effect of competitor type (fewer fixations to rhyme than to onset competitors: $\beta = -0.741$, $t = -3.86$, $p < 0.001$) and no effects of noise or noise position ($t < 1$ for each noise condition relative to baseline and for the comparison between the two noise conditions, and $t < 1$ for the modulation of any noise-position effect by competitor type). But there were significant interactions of the two factors (relative to the baseline, more fixations to rhyme competitors and fewer fixations to onset competitors in both the word-onset noise condition, $\beta = -0.297$, $t = 2.59$, $p < 0.02$, and the word-medial noise condition, $\beta = -0.274$, $t = 2.46$, $p < 0.02$).

B. Critical time-window analyses

The overall analyses thus suggest that, as noise (in either position) is added to the carrier sentences, fixations to onset competitors decrease and fixations to rhyme competitors increase. As noted in Sec. I, however, the critical predictions concern the 200–400 ms region. We expected [based on the findings of Allopenna *et al.* (1998) and McQueen and Viebahn (2007)] preferential looks in the baseline condition in this time window to onset competitors, but not to rhyme competitors. Figure 3 shows that this was the case. A change in this pattern in the noise conditions would reveal that the noise manipulation had affected the earliest moments of word recognition. Figure 3 also shows that this was the case. Participants in the noise groups looked in this time window more at the rhyme competitors and less at the onset competitors than participants in the baseline group.

We carried out another 2×2 repeated-measures ANOVA, but now on the fixation ratio data in the 200–400 ms region. These analyses revealed a main effect of competitor type [$F_1(1,53) = 42.50$, $p < 0.001$; $F_2(1,73) = 15.36$, $p < 0.001$], no main effect of noise (F_1 and $F_2 < 1$), and a noise by competitor-type interaction [$F_1(1,53) = 7.60$, $p = 0.008$; $F_2(1,73) = 3.80$, $p = 0.07$]. Thus the presence or absence of radio noise influenced participants' fixation behavior. A further 2×2 repeated-measures ANOVA with noise position (word-onset vs word-medial) as between subject factor revealed no noise position by competitor type interaction [$F_1(1,32) = 0.13$, $p > 0.1$; $F_2(1,48) = 0.20$, $p > 0.1$]. The position of the radio noise (onset or medial) therefore had no significant effect.

LMERs of logit competitor fixation proportions in the critical region, like those in the overall analysis, confirmed these findings. There was an effect of competitor type ($\beta = -0.941$, $t = -4.67$, $p < 0.001$), no effects of noise (relative to baseline, for word-onset noise: $\beta = -0.266$, $t = -1.38$, $p > 0.15$, and for word-medial noise: $\beta = -0.323$, $t = -1.73$, $p = 0.08$). Critically, there were significant interactions of competitor type and noise (relative to the baseline: word-onset noise, $\beta = 0.508$, $t = 2.58$, $p < 0.02$; word-medial noise, $\beta = 0.507$, $t = 2.66$, $p < 0.009$). In a direct comparison of the two noise positions, there was no difference between word-initial and word-medial noise ($t < 1$), and this was not modulated by competitor type ($t < 1$).

Pairwise comparisons were then carried out on the ratio data for each competitor type in each condition in the 200–400 ms region (see also Fig. 3). There were significantly more looks to onset competitors than to distractors in the baseline condition [mean ratio 0.66, by participant range 0.51–0.80, $t_1(20) = 9.89$, $p < 0.001$; $t_2(24) = 4.90$, $p < 0.001$]. Although there were numerically fewer looks to onset competitors in the noise conditions, both effects remained statistically significant [word-onset-noise: mean ratio 0.62, range 0.49–0.74, $t_1(15) = 6.37$, $p < 0.001$; $t_2(24) = 2.18$, $p = 0.039$; word-medial-noise: mean ratio

J. M. McQueen and F. Huettig: Flexibility in spoken-word recognition

0.60, range 0.26–0.72, $t_1(17) = 2.48$, $p = 0.024$; $t_2(24) = 2.46$, $p = 0.022$]. There were no more looks to rhyme competitors than to distractors in the baseline condition (mean ratio 0.49, range 0.28–0.65, t_1 and $t_2 < 1$). Although there was a numerical increase in fixations to rhyme competitors in the noise conditions, neither effect was statistically significant [word-onset-noise: mean ratio 0.54, range 0.35–0.72, $t_1(15) = 1.46$, $p < 0.1$; $t_2(24) = 1.73$, $p = 0.097$; word-medial-noise: mean ratio 0.54, range 0.34–0.74, $t_1(17) = 1.39$, $p > 0.1$; $t_2 < 1$]. These tests suggest that the critical interaction of competitor type by noise condition is not driven by a change in fixation behavior for either type of competitor alone (i.e., it is not the case that there is only a decrease in fixations for onset competitors or only an increase in fixations for rhyme competitors). Instead, the interaction reflects the joint effect of small changes for both types of competitor. In other words, as noise is added, looks to onset competitors go down, as looks to rhyme competitors go up.

C. Additional time-window analyses

ANOVAs on fixation ratios and LMERs on competitor fixation logits (like those for the 200–400 ms region) then examined performance in three further 200 ms regions: 0–200, 400–600, and 600–800 ms. In the earliest region, there were no significant effects of competitor type, noise, or their interaction in either type of analysis. The differences between conditions visible in this region in Fig. 1 are thus spurious, presumably because they do not yet substantially reflect behavior driven by the acoustics of the critical words (Matin *et al.*, 1993; Saslow, 1967). In the later regions, there was an advantage for onset over rhyme competitors that weakened to some extent later in time [400–600 ms region: $F_1(1,53) = 40.32$, $p < 0.001$, $F_2(1,73) = 9.94$, $p = 0.002$, LMER $\beta = -0.772$, $t = -3.47$, $p < 0.001$; 600–800 ms region: $F_1(1,53) = 19.55$, $p < 0.001$, $F_2(1,73) = 3.41$, $p = 0.07$, LMER $\beta = -0.510$, $t = -2.02$, $p < 0.05$]. But there were no significant effects of noise or of noise position, and, importantly, no significant interactions between noise and competitor type. The critical interaction thus appears to be confined to the 200–400 ms region.

IV. DISCUSSION

This experiment shows that the perceptual weight assigned to acoustic information during spoken-word recognition changes as a function of the context in which that information is heard. Dutch listeners looked at pictures of phonological competitors of spoken words, presented in sentences recorded in high fidelity. In line with previous studies (Allopenna *et al.*, 1998; McQueen and Viebahn, 2007), adult listeners looked much more at competitors with overlapping onset sounds in their names (e.g., a picture of a crocodile [*krokodil*] as they heard *krokus*) than at rhyming competitors (e.g., a hammer [*hamer*] as they heard *kamer*). This tendency remained, but became weaker when noises obliterated some speech sounds elsewhere in the spoken sentences. The weight assigned to the information at the beginning of spoken words thus appeared to be changed in the adverse-

listening situation of hearing speech masked by occasional bursts of noise, as occurs when we hear a badly tuned AM radio.

We suggest that the presence of occasional noises in the sentences changed the certainty that listeners had about the information in the words they were hearing. This had two effects: The noises made the listeners less confident about what the word onsets were, and also about what the onsets were *not*. As they heard the beginning of *krokus*, for example, they became less sure, relative to the no-noise baseline, that they were hearing the beginning of *krokus* or *krokodil*, so they looked less at the crocodile. Furthermore, as they heard *kamer*, for example, they became less sure that they had heard *kamer*, but also less sure that they had not heard *hamer*, so they looked more at the hammer.

The effects of adding intermittent noises were found only 200–400 ms after onset of the critical words. This suggests that the perceptual adjustment which took place in the noise conditions affected the earliest moments of word recognition, as would be expected if the adjustment is an anticipatory process (i.e., such that it has an immediate influence on the way the first sounds of the words are processed). As predicted, the effect of the noise weakened later in time, because, as processing unfolds, it should become clear that the initial sound of the critical word was not replaced by noise. The acoustic-phonetic information about the word's initial sound is thus available to influence word recognition, and indeed appears to do so. This also explains why in the noise conditions, even in the 200–400 ms region, there are more looks to the competitors than to the distractors: The acoustic-phonetic information supports the competitors but not the distractors. The effects of the noise manipulation are early, small, and fleeting because the acoustic evidence at the beginning of the word is undistorted, and hence is available to influence word recognition.

There was no significant effect of the manipulation of the position of the noise. Relative to the baseline condition, the changes in fixation behavior were the same when the noises were always in word-onset position as when they were always in word-medial position. This suggests that the adjustment made to speech with radio noise is a general one, rather than one specific to noise position. The speech-recognition system thus appears to be able to detect that some phonemes were being masked by noise, but not where those phonemes were. It has long been known that listeners have difficulty in detecting the location of clicks (Fodor and Bever, 1965), and that listeners are susceptible to the phonemic restoration illusion, where noises (e.g., coughs) replace sounds in words (Warren, 1970). Temporal localization of extraneous sounds in speech thus appears to be difficult. Another reason why the listeners in the present experiment appear to have been unable to detect the location of the noises may be due to the nature of radio interference outside the laboratory. In the real world, interference is very unlikely always to be in onset position, so the current participants are not likely to have had prior experience that encouraged them to learn how to make position-specific adjustments. But they probably have experienced random obliteration of pieces of speech before. This calls for the ability to make a global

adjustment when speech with occasional noise interference is heard, and it is this kind of flexibility that the present experiment is measuring.

Previous research has indicated that listeners can adjust how acoustic information is interpreted. Listeners use lexical knowledge, for example, to tune in to speech sounds that have been spoken in an unfamiliar way (Norris *et al.*, 2003; McQueen *et al.*, 2006). Such retuning of perception helps listeners deal with the variability of speech. The present study reveals a new kind of perceptual flexibility. Listeners appear not only able to adjust how they interpret acoustic information, they can also adjust the weight that they assign to the same acoustic information in different listening situations. Brouwer *et al.* (2011) have recently demonstrated a similar kind of listener flexibility using eye tracking. The degree to which different words compete during spoken-word recognition appears to change depending on whether the listener is hearing speech with phonological reductions (e.g., "puter" for "computer") or fully articulated speech with no such reductions. These kinds of plasticity allow listeners to cope better with speech variability.

Importantly, the noises never occurred on the critical words themselves. This means that the change in the recognition process observed in the noise conditions is a response to the increased probability of noise occurring, and cannot be a response to the acoustic consequences of the noise itself. This sensitivity to changing probabilities is consistent with the Bayesian account of speech recognition proposed in the Shortlist B model (Norris and McQueen, 2008). According to this account, word recognition is modulated both by prior probabilities (e.g., high-frequency words are recognized more quickly than low-frequency words) and by word likelihoods (the probability that particular acoustic–phonetic evidence would occur given that a certain word was spoken). Word likelihoods could change depending on the probability of noise in a given context [e.g., p($krok…$ in input | $krokodil$ spoken) gets smaller with noise, and p($kam…$ in input | $hamer$ spoken) gets larger].

Computational work would be required to specify how these changes in word likelihoods could be implemented in Shortlist B. The model's Bayesian approach nevertheless offers a promising framework for explaining the current results. If perception is based on Bayesian inference [see, e.g., Feldman *et al.* (2009), and Yuille and Kersten (2006), for domains other than spoken-word recognition], then perception is predicted to change across environments to ensure that it is optimal in any given situation. Spoken-word recognition should thus be optimal both when speech is disrupted by radio noise and when it is not. Intermittent noise makes speech information less reliable, and the probabilistic computation in Shortlist B could capture exactly that.

It is important to note, however, that the present results may also be compatible with other models of spoken-word recognition. In the Neighborhood Activation Model (Luce and Pisoni, 1998), for example, the weighting of the effect of bottom-up acoustic–phonetic information in the neighborhood computation could be altered in different noise conditions. Similarly, the degree of bottom-up activation of word nodes by phoneme nodes in TRACE (McClelland and Elman, 1986) could change in the presence of noise disruptions. Alternatively, the presence of intermittent noise could cause changes in the k parameter in TRACE, a parameter that governs the relationship between lexical activation and response probability in the Luce choice rule at the model's decision stage. Mirman *et al.* (2011) have shown that changes in TRACE's k parameter can indeed simulate differences in fixation behavior in spoken-word recognition experiments with aphasic patients. When the value of k is low, more rhyme-competitor fixations are predicted (as observed in Broca's aphasics). When the value of k is high, more onset-competitor fixations are predicted (as observed in Wernicke's aphasics). Further research will be required to determine how the present findings can best be modeled. These findings nevertheless suggest that, in all future models, a flexibility needs to be added through which the presence of distortions in the current input can alter the way acoustic–phonetic information influences spoken-word recognition.

At one level, therefore, this investigation confirms that the same word-recognition process operates with and without intermittent noise interference: There is always optimal and incremental evaluation of acoustic–phonetic information. At a more detailed level, however, this study shows that changing the probability that a word might be distorted by radio noise does change how that word is recognized. Exactly the same information is treated in one way when intermittent noises are present elsewhere in the input, and in another way in the high-fidelity listening situation that is normally tested in the perceptual laboratory.

The present findings therefore do not completely undermine previous conclusions based on findings with high-quality speech. The core process of optimal interpretation appears to apply across listening situations. But the present findings do show that investigations restricted to careful laboratory speech will limit understanding of speech processing. One possibility is that research on word recognition in noise-free environments could lead to distorted views on what constitutes default recognition processes (e.g., that onset-overlap competitors are always going to be stronger than rhyme-overlap competitors). It is certainly the case that such investigations would fail to detect a key property of the word-recognition system: Its flexibility in response to changing probabilities in the speech signal, across good and adverse listening conditions.

ACKNOWLEDGMENTS

APPENDIX

See Table I.

TABLE I. Critical stimuli.

Overlap	Spoken word	Translation	Picture competitor	Translation
Onset	agenda	agenda	agent	policeman
	blikopener	can opener	bliksem	lightening
	circus	circus	cirkel	circle
	kompas	compass	computer	computer
	diabolo	diabolo	diamant	diamond
	duiventil	pigeon house	duivel	devil
	hagelslag	chocolate sprinkles	hagedis	lizard
	kers	cherry	kerstboom	christmas tree
	krans	wreath	krant	newspaper
	krokus	crocus	krokodil	crocodile
	matras	mattress	matroos	sailor
	paraplu	umbrella	parachute	parachute
	parkeermeter	parking meter	parkiet	parrot
	paspoort	passport	paspop	tailor's dummy
	pepernoot	biscuit	paperclip	paper clip
	pyjama	pyjamas	piano	piano
	schilder	painter	schildpad	turtle
	synagoge	synagogue	sinaasappel	orange
	stemvork	tuning fork	stempel	stamp
	steiger	pier	stijgbeugel	stirrup
	strippenkaart	ticket strip	strik	knot
	telefooncel	call box	telescoop	telescope
	telefoon	telephone	televisie	television
	thermoskan	thermos	thermometer	thermometer
	vliegenmepper	fly swatter	vlieger	pilot
Rhyme	riool	sewer	viool	violin
	kabel	cable	sabel	sword
	kamer	room	hamer	hammer
	honing	honey	koning	king
	wafel	waffle	tafel	table
	suiker	sugar	duiker	diver
	zegel	stamp	kegel	ninepin
	borst	chest	worst	sausage
	varen	fern	garen	thread
	laken	sheet	baken	beacon
	gids	guide	rits	zip
	laars	boot	kaars	candle
	haard	fireplace	baard	beard
	paard	horse	taart	cake
	golf	wave	wolf	wolf
	dolk	dagger	wolk	cloud
	cent	cent	tent	tent
	graan	grain	kraan	tap
	mast	mast	kast	cupboard
	hond	dog	mond	mouth
	lans	lance	gans	goose
	mand	basket	tand	tooth
	park	park	hark	rake
	kruik	jug	pruik	wig
	draad	thread	graat	fish bone

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

Anderson, J. R. (**1990**). *The Adaptive Character of Thought* (Erlbaum, Hillsdale, NJ).

Baayen, R. H., Davidson, D. J., and Bates, D. M. (**2008**). "Mixed-effects modelling with crossed random effects for subjects and items," J. Mem. Lang. **59**, 390–412.

Barr, D. J. (**2008**). "Analyzing 'visual world' eye tracking data using multilevel logistic regression," J. Mem. Lang. **59**, 457–474.

Boersma, B., and Weenink, D. (**2005**). PRAAT: Doing phonetics by computer (Version 4.3.14). Retrieved May 26, 2005, from http://www.praat.org.

Brouwer, S., Mitterer, H., and Huettig, F. (**2011**). "Speech reductions change the dynamics of competition during spoken word recognition," Lang. Cognit. Processes, doi:10.1080/01690965.2011.555268.

Cherry, E. C. (**1953**). "Some experiments on the recognition of speech, with one and with two ears," J. Acoust. Soc.Am. **25**, 975–979.

Connine, C. M., Blasko, D. G., and Titone, D. (**1993**). "Do the beginnings of spoken words have a special status in auditory word recognition?," J. Mem. Lang. **32**, 193–210.

Desroches, A. S., Joanisse, M. F., and Robertson, E. K. (**2006**). "Specific phonological impairments in dyslexia revealed by eyetracking," Cognition **100**, B32–B42.

Feldman, N. H., Griffiths, T. L., and Morgan, J. L. (**2009**). "The influence of categories on perception: Explaining the perceptual magnet effect as optimal statistical inference," Psychol. Rev. **116**, 752–782.

Fodor, J. A., and Bever, T. G. (**1965**). "The psychological reality of linguistic segments," J. Verbal Learn. Verbal Behav. **4**, 414–420.

Frauenfelder, U. H., Scholten, M., and Content, A. (**2001**). "Bottom-up inhibition in lexical selection: Phonological mismatch effects in spoken word recognition," Lang. Cognit. Processes **16**, 583–607.

Huettig, F., and Altmann, G. T. M. (**2005**). "Word meaning and the control of eye fixation: Semantic competitor effects and the visual world paradigm," Cognition **96**, B23–B32.

Huettig, F., and McQueen, J. M. (**2007**). "The tug of war between phonological, semantic, and shape information in language-mediated visual search," J. Mem. Lang. **57**, 460–482.

Luce, P. A., and Pisoni, D. B. (**1998**). "Recognizing spoken words: The neighborhood activation model," Ear Hear. **19**, 1–36.

Magnuson, J. S., Tanenhaus, M. K., Aslin, R. N., and Dahan, D. (**2003**). "The time course of spoken word recognition and learning: Studies with artificial lexicons," J. Exp. Psychol. Gen. **132**, 202–227.

Marslen-Wilson, W., Moss, H. E., and van Halen, S. (**1996**). "Perceptual distance and competition in lexical access," J. Exp. Psychol. Hum. Percept. Perform. **22**, 1376–1392.

Marslen-Wilson, W., and Welsh, A. (**1978**). "Processing interactions and lexical access during word recognition in continuous speech," Cogn. Psychol. **10**, 29–63.

Marslen-Wilson, W., and Zwitserlood, P. (**1989**). "Accessing spoken words: The importance of word onsets," J. Exp. Psychol. Hum. Percept. Perform. **15**, 576–585.

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

McClelland, J. L., and Elman, J. L. (**1986**). "The TRACE model of speech perception," Cogn. Psychol. **18**, 1–86.

McQueen, J. M., Cutler, A., and Norris, D. (**2006**). "Phonological abstraction in the mental lexicon," Cogn. Sci. **30**, 1113–1126.

McQueen, J. M., and Viebahn, M. (**2007**). "Tracking recognition of spoken words by tracking looks to printed words," Q. J. Exp. Psychol. **60**, 661–671.

Miller, G. A., and Licklider, J. C. R. (**1950**). "The intelligibility of interrupted speech," J. Acoust. Soc. Am. **22**, 167–173.

Mirman, D., Yee, E., Blumstein, S., and Magnuson, J. S. (**2011**). "Theories of spoken word recognition deficits in aphasia: Evidence from eye-tracking and computational modeling," Brain Lang. **117**, 53–68.

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

Norris, D., McQueen, J. M., and Cutler, A. (**2003**). "Perceptual learning in speech," Cogn. Psychol. **47**, 204–238.

Norris, D., and McQueen, J. M. (**2008**). "Shortlist B: A Bayesian model of continuous speech recognition," Psychol. Rev. **115**, 357–395.

Payton, K. L., Uchanski, R. M., and Braida, L. D. (**1994**). "Intelligibility of conversational and clear speech in noise and reverberation for listeners with normal and impaired hearing," J. Acoust. Soc. Am. **95**, 1581–1592.

Reinisch, E., Jesse, A., and McQueen, J. M. (**2010**). "Early use of phonetic information in spoken word recognition: Lexical stress drives eye-movements immediately," Q. J. Exp. Psychol. **63**, 772–783.

Saslow, M. G. (**1967**). "Latency for saccadic eye movement," J. Opt. Soc. Am. **57**, 1030–1033.

Warren, R. M. (**1970**). "Perceptual restoration of missing speech sounds," Science **167**, 392–393

Yuille, A., and Kersten, D. (**2006**). "Vision as Bayesian inference: Analysis by synthesis?," Trends Cogn. Sci. **10**, 301–308.

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