

A contingent speech technique in eye movement research on reading

ALBRECHT W. INHOFF and CYNTHIA CONNINE
State University of New York, Binghamton, New York

and

RALPH RADACH
Technical University at Aachen, Aachen, Germany

A novel eye-movement-contingent method is presented. It builds on and extends established eye-movement-contingent visual display change methods in that it uses movements of the eyes to control the presentation of acoustic information during sentence reading. In one implementation, an irrelevant spoken word is presented when the eyes cross a predetermined spatial boundary before they move on to a selected visual target word. The relationship between the spoken word and the visual target is manipulated, and the pattern of interference, caused by the presentation of the spoken word, is used to determine the nature and time course of activated representations. Results from three recently completed experiments in which the technique was used show that a word's phonological code remains active after it has been read and that the activated code has speech-like properties.

Readers need to move their eyes to obtain useful linguistic information from different segments of text. The two components of eye movements in reading, saccades and fixations, are sensitive to a wide range of perceptual and linguistic processes (see Radach & Kennedy, in press, and Rayner, 1998, for recent reviews), and their continuous measurement provides distinct methodological advantages over other measures of perceptual and linguistic processing. Eye movement measures can be obtained under natural task conditions, they provide information about the short- and long-term time course of information usage, and they obviate the execution and measurement of responses, such as manual buttonpresses, that are extraneous to the reading task.

It is largely for these reasons that eye movement measures are increasingly used in the study of reading and other cognitive processes. Consequently, a large part of the eye movement research community currently consists of researchers studying different aspects of written language processing (see Kennedy, Radach, Heller, & Pynte, 2000, for a discussion of the state of the art in this area). The vast majority of these researchers employ eye movements as indicators of the various levels and modules of cognitive processes. As we noted in a recent summary of the state of measurement-related discussions (Inhoff & Radach,

1998), eye movements can also be used to compute a range of processing measures that can be designed to meet specific experimental needs (Inhoff & Weger, in press).

The expanded use of eye movement methods also calls for methodological innovation and for new tools in the study of cognitive processes. The present paper focuses on the most sophisticated and methodologically complex form of eye movement experimentation, the use of saccades for the time-locked manipulation of available information. We will first briefly consider some central methodological aspects of this type of research and then describe a new, innovative procedure, the *contingent speech technique*.

One methodological limitation in studies using oculomotor measures is that the researcher is not in control of the eye movements that a participant makes in response to a particular experimental manipulation. For instance, in a study of perceptual processes during reading, the experimenter cannot determine how much visual information is available during a fixation, for how long a studied segment of text is being viewed, and what type of information is available during a fixation.

In response to this, Rayner (1975) and McConkie and Rayner (1975) introduced the eye-movement-contingent display change technique. In this technique, movements of the eyes are monitored and used to control when and where useful information is available on a computer screen. In one version, Rayner's (1975) boundary technique, the experimenter selects a spatial location within a to-be-read sentence, the display change boundary. Eye position is monitored, and the saccade that moves the eyes to the right of (or at) the boundary initiates a change of available information. In the original study, the display change was

This research was supported by NSF Grant 0002024 and by the Center for Cognitive and Psycholinguistic Sciences at the State University of New York at Binghamton. We appreciate Steffi Lemmer's help in the laboratory and the constructive comments of three reviewers. Please address correspondence to A. W. Inhoff, Department of Psychology, State University of New York at Binghamton, Binghamton, NY 13902 (e-mail: inhoff@binghamton.edu).

used to control the type of information that a reader could obtain from a selected sentence location. At the onset of sentence reading, the (target) word at that location was occupied by a length-matched string of letters that differed from the target in controlled ways. The boundary was immediately to the left of the target, and its eventual crossing during sentence reading replaced the letter string with the target. Time spent viewing the target was then used to determine the use of information from the previously available letter string (see McConkie, Wolverson, & Zola, 1984, for an introduction to key technical and methodological issues).

More recent versions, such as Sereno and Rayner's (1992) fast priming technique, have extended experimental control via precise manipulations of the spatial and temporal properties of available information. Similar to the original boundary study, all the words of a sentence were visible throughout sentence reading, except for a target word, each letter of which was replaced by a visually dissimilar letter at the onset of sentence reading. However, in contrast to Rayner (1975), the relationship between the properties of the initially (parafoveally) visible letter string at the target location and the target word was not manipulated. Instead, the movement of the eyes across the boundary replaced the letter string briefly with a prime word, and the relationship between the prime and the subsequently presented target word was manipulated.

Measurements of target gazes—that is, the cumulated duration of all target fixations until another word was fixated—as a function of prime type indicated that readers obtained useful phonological and semantic information from primes that were presented for less than 40 msec after the onset of the first fixation at the target location. Furthermore, effects of homophony emerged for shorter prime durations than did effects of word meaning. Interestingly, effects of homophonic primes ceased to exist for prime durations of 38 msec and more, suggesting that the effects of homophony emerged only when the prime's identity was not fully (consciously) established.

Despite their power and flexibility, the use of eye-movement-contingent visual display changes remains, to some extent, limited by eye movement properties. For instance, the duration of primes presented at fixation onset must be relatively short, because the eyes may leave the target location before the target word is presented when the prime duration is relatively long. The method also requires careful visual control conditions, because the oculomotor system not only implements the experimental manipulation, but also responds to it. Specifically, an eye-movement-contingent display change during a fixation, designed to control processing of linguistic information, may also cause an unrelated time-locked inhibition of a programmed saccade (Reingold & Stampe, 2000).

We recently developed a novel boundary technique, the eye-movement-contingent presentation of speech technique, that circumvents some of the shortcomings of visual display change manipulations. The key event in our technique is the presentation of an *auditory* stimulus when

the eyes cross a spatial boundary during sentence reading, so that the visibility of a visual target word and the onset of the auditory stimulus are synchronized. In our current utilization of the technique, the auditory event is a word whose phonological properties are manipulated relative to the visual target. The experimental measure consists of gaze durations for the target word and for subsequent words in the sentence that are examined as a function of spoken word type.

The impetus for the development of this technique was twofold: to chart the time course of phonological code usage during sentence reading and to determine the nature of a word's phonological code. Although there is a relatively large body of evidence indicating use of phonological information during visual word recognition (Frost, 1998; see Starr & Fleming, 2001, for a different perspective), relatively little is known about the time course of phonological code usage—that is, whether and for how long an activated phonological form survives the recognition of the corresponding word. The nature of the phonological code has also remained elusive. The default assumption is that activation of phonological representations involves some form of grapheme-to-phoneme conversion or the retrieval of a corresponding form (Frost, 1998). Others, however, have argued that the phonological code is akin to inner speech (e.g., Hochberg, 1976), and there is recent evidence suggesting that the phonological code may have speech-like properties (Birch, Pollatsek, & Kingston, 1998).

In Experiments 1 and 2, the contingent speech technique was used to synchronize the presentation of a visual target word and its spoken companion word during sentence reading. The phonological form of the spoken word was identical, similar, or dissimilar to the phonological form of the visual target, and neither similar nor dissimilar spoken words fit prior and subsequent sentence context. Phonological properties of the spoken word were expected to influence target reading if printed and spoken words activated identical or structurally related phonological representations. Under the present conditions, a similar spoken word should hamper visual target word reading, relative to a dissimilar and an identical spoken word, since its phonological representation should compete with the phonological representation of the visual target for recognition. Moreover, if the phonological representations of the visual target and its spoken companion word remained active after the two words were identified, deleterious effects of the similar spoken word should extend beyond target reading.

EXPERIMENT 1

Method

Participants. Thirty undergraduate or graduate students with no history of reading disabilities participated in the study for course credit or payment.

Materials. The reading material consisted of 56 declarative sentences, 42 experimental sentences, and 14 fillers, each of which contained 8–13 words. The structure, length, and content of these sen-

tences were similar to the experimental stimuli used in other reading research that has used eye-movement-contingent visual display changes to study phonological representations in reading. Each experimental sentence contained a visual target word, a noun, which occurred near the middle of the sentence.

Three to-be-spoken content words (nouns or adjectives) were selected for each visual target: one that was identical in phonological form and meaning to the visual target, one that was similar in phonological form and dissimilar in meaning, and one that was dissimilar in phonological form and in meaning. For instance, the visual target plate was presented with the spoken word PLATE in the identical condition, with PLACE in the similar condition and with HORSE in the dissimilar condition (see Figure 1). All the words were spoken without a distinguishing local accent by a male native speaker of American English. Speech-editing software (Soundforge 4.5) was used to record spoken words in 16-bit stereo at 20500 Hz, to remove onset pauses from the spoken word file and to set loudness levels to approximately 70 db. Articulation durations ranged from approximately 350 to 700 msec, with a mean of 546 msec for identical spoken words and of 534 msec for similar and dissimilar spoken words.

Apparatus. All text was displayed on a gray background on a 21-in. flat-screen Liyama Vision Master 510 monitor with .28-mm dot pitch with a $1,028 \times 720$ pixel resolution. Black Courier type font was used so that each character occupied the same horizontal area

of text, with a maximum of 12 horizontal pixels per character. The distance between the readers' eyes and the monitor was set at approximately 85 cm. However, since head movements were not restrained the distance was somewhat variable. At a typical distance of 85 cm, each letter of text subtended approximately 0.44° of visual angle laterally.

Eye movements were recorded via a head-mounted EyeLink pupil-tracking system. Viewing was binocular, but eye movements were recorded from the right eye only. A high-speed video camera, used for the recording of pupil position, was positioned underneath the monitored eye and was held in place by head-mounted gear. A second camera, mounted at the front of the headband, monitored head position. Both cameras sampled at a rate of 250 Hz throughout sentence reading, and the combined monitoring of pupil and head position yielded an absolute visual resolution in the order of better than 0.5° (approximately the width of one letter) and a relative resolution of better than 0.25° even in the presence of minor head movements.

Acoustic signal presentation was controlled by dedicated hardware (SoundblasterPro). Spoken words were presented via two speakers (Altec Lansing) positioned to the right and left of the screen, approximately 80 cm in front of the target. Speakers, rather than headphones, were selected to avoid additional attachments to the reader's head.

A DOS-based software package (the EasyProgrammingKit, EPK), developed at the Technical University of Aachen (Craemer &

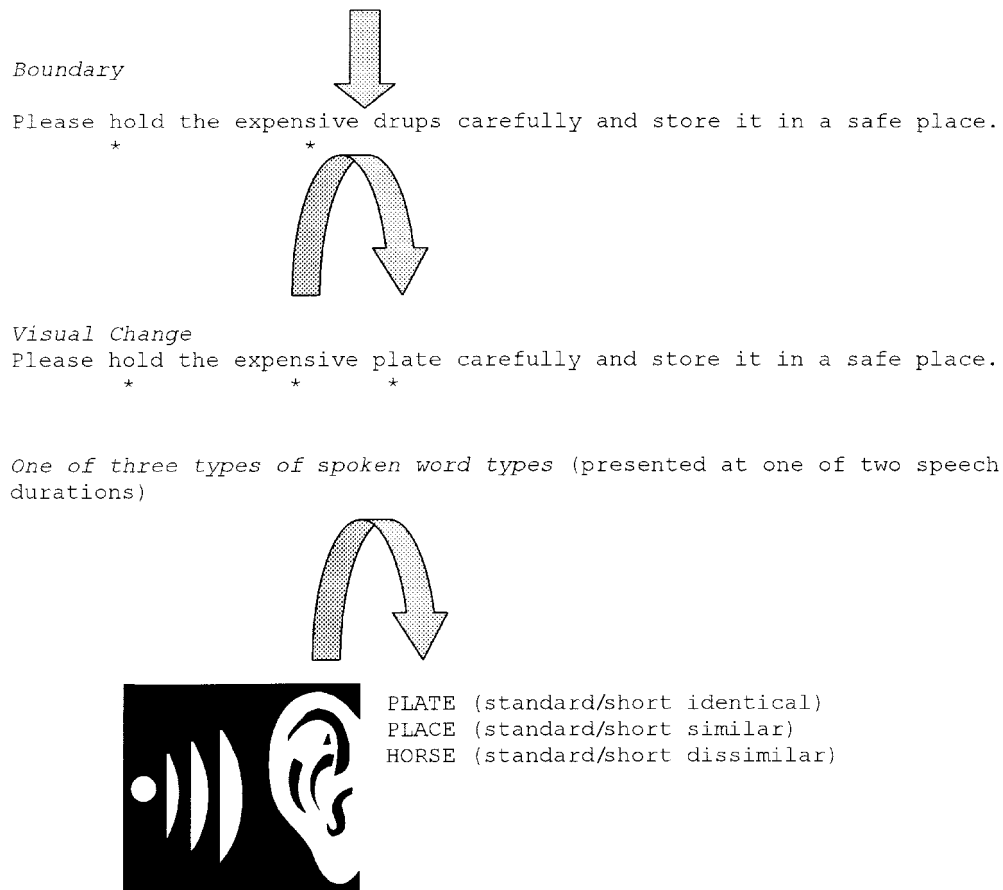


Figure 1. A sample sentence along with pretarget and target fixations, indicated by asterisks. The arrow shown for the top line reveals the sentence boundary. The saccade across the boundary, shown in the second line, revealed the visual target word. In addition, as is shown in the third line, it also led to the presentation of a spoken word that was identical, similar, or dissimilar to the visual target. Two articulation rates were used: standard in Experiment 1, and short (compressed speech) in Experiment 2.

Radach, 2002), was used to implement eye-movement-contingent manipulations of visual text and the contingent presentation of speech. In this software package, all information necessary for running complex eye movement experiments is implemented in a simple script language. It incorporates a number of C-type commands and hardware-specific procedures provided by SR research as part of the software for the EyeLink tracker.

Procedure. The experimental session began with a one-dimensional horizontal calibration of the eye-tracking system, which required be-

tween 3 and 5 min. This involved the sequential presentation of five fixation markers at one of three horizontal screen locations (left, center, or right) at the vertical midline of the screen. The spatial ordering of fixation markers was randomized, and each marker remained visible for 1 sec. A calibration was considered successful when the fixations of a left-side, center, and right-side fixation marker were within less than 0.5° of the marker. After that, the participants were asked to read a sequence of unrelated sentences for comprehension while their eye movements were monitored. One sentence

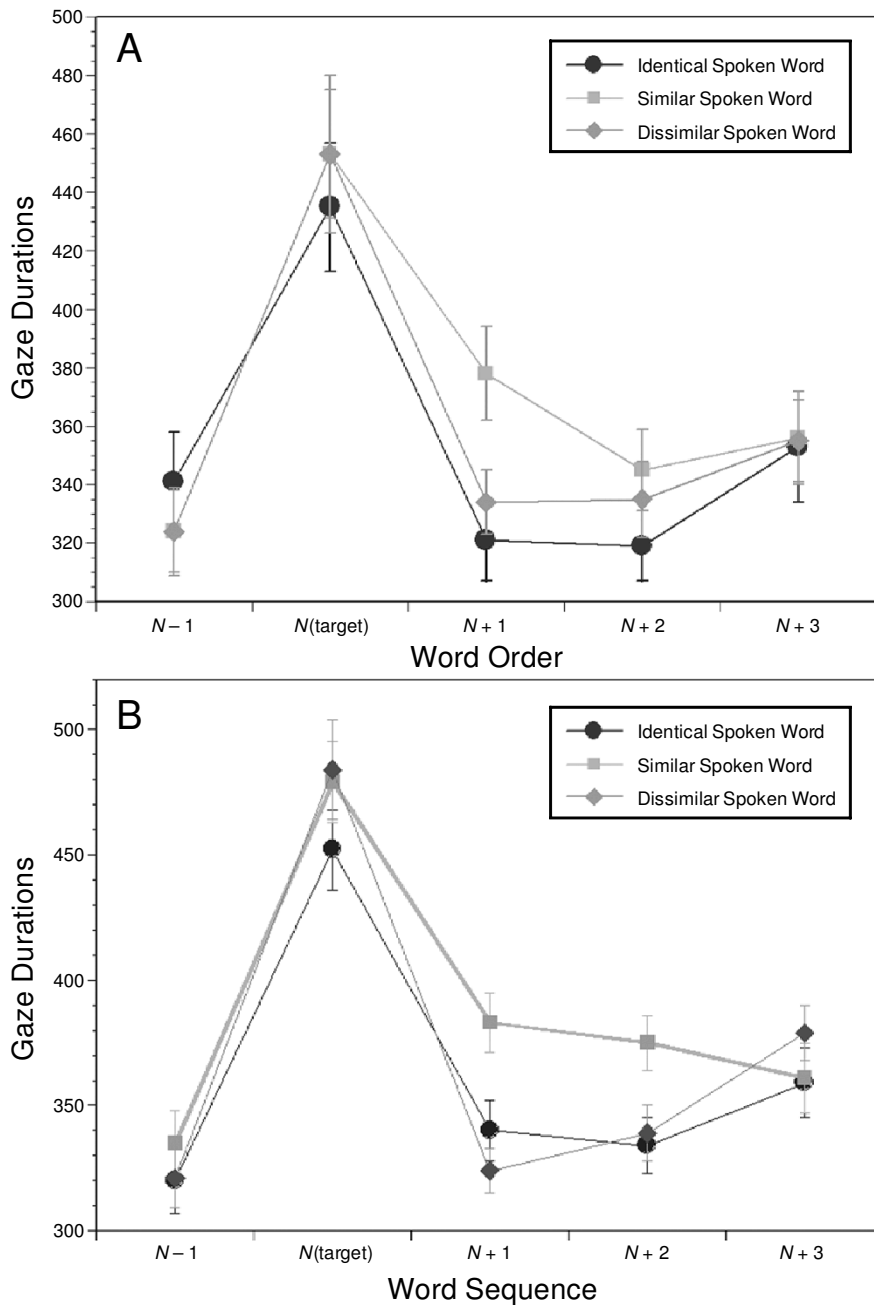


Figure 2. Gaze durations as a function of spoken word type for the three spoken word conditions in Experiment 1 (A) and Experiment 2 (B). Gazes are shown for the pretarget word, before the eye-movement-contingent presentation of speech occurred, for the target, and for the following three posttarget words; the third word generally coincided with the sentence ending.

was shown at a time at the horizontal screen location of previously shown fixation markers, and the onset and offset of the sentence were controlled by the participant by pressing the space bar on a computer keyboard. The words of a sentence were continuously visible during sentence reading, except for the visual target of an experimental sentence and the visual filler target of a filler sentence. A placeholder, consisting of an orthographically regular pseudoword, occupied the target and filler target location at the onset of sentence reading. The phonological form of the placeholder was unrelated to the phonological form of the corresponding visual target and to the phonological forms of a target's three spoken companion words. The placeholder was changed to the target and the spoken companion word was presented when the eyes moved to the right of the blank space preceding the target location (the target boundary).

The spoken word file was preloaded at the onset of a trial, and release of the acoustic signal occurred within 4–6 msec after the eyes reached the target boundary. Implementation of the visual display change generally took another 6–10 msec. Specifically, the physical imaging of the camera input took 2–4 msec; processing of this image then required another 4 msec, data smoothing (to obtain a more stable signal) required 4 msec, and the transfer of location data to the display computer required 1 msec. Subsequent implementation of the display change on the CRT also had to wait for the next refresh cycle. At a CRT refresh rate of 160 Hz, this additional wait state ranged from 1 to 6 msec, yielding a visual display change implementation lag between 13 and 18 msec. However, this delay is unlikely to distort experimental effects. We have found in other work (Inhoff, Radach, Eiter, & Juhasz, in press) that the effects of eye-movement-contingent visual display changes were quite similar when implemented with the EyeLink system and with a somewhat faster Dual-Purkinje system, where typical implementation delays ranged from 4 to 10 msec, and when there were small delays in the implementation of visual display changes of 10 msec or less at fixation onset (Inhoff, Starr, Liu, & Wang, 1998).

Even if the somewhat slower implementation of the display change had some spurious effect on eye movements, this should not impede the interpretation of experimental effects, since the same words were viewed under identical viewing conditions in the three spoken word conditions. The reading of every sentence was followed by a new calibration and the reading of another sentence.

The readers were told that either a word or white noise would be presented over speakers at some point in time during sentence reading and that there would not be any testing for it. To ensure that the participants focused on sentence meaning, they were asked to paraphrase a previously read sentence approximately every 10th sentence. The readers paraphrased virtually all sentences correctly (>98%), and they never reported the spoken word in lieu of the visual target.

Design and Measurement. Three lists were used so that each visual target was paired with each spoken word type and so that each list contained the same number of sentences in each one of the three experimental conditions. The spoken words used in the similar condition of one list were paired with a different-sounding visual target on another list, to create the dissimilar condition. To determine effects of speech type on target word and posttarget reading, we computed a word's gaze duration, provided it exceeded 50 msec. Gaze durations consisted of the cumulated duration of all fixations on a word until another word was fixated. Saccade durations—that is, the time spent moving eyes within a word—were not included in the measure. The gaze durations of five consecutive (critical) words of each target sentence were computed, starting with the pretarget word ($N - 1$) and the target (N) and ending with the third posttarget word ($N + 3$).

Results

Figure 2A shows gaze durations for the five critical words as a function of spoken word type.

As can be seen, the presentation of the spoken companion word immediately increased the visual target's gaze duration over pretarget gaze durations. The increase was followed by a rapid drop in gaze durations for posttarget words. This occurred irrespective of word type and was expressed in a reliable effect of word location [$F(4,116) = 22.88, p < .01$]. Presentation of white noise (on filler trials) showed a similar increase, with gaze durations of 258, 423, 307, 330, and 373 msec for positions $N - 1$ to $N + 3$, respectively, suggesting that the sudden onset of an acoustic signal briefly disrupted sentence reading irrespective of acoustic signal type. All the spoken word conditions and filler trials also revealed a general increase in gaze durations for $N + 2$ and $N + 3$ words. These words were close to or at the sentence ending, and the increase in viewing duration may be related to previously reported sentence *wrap-up* effects (Just & Carpenter, 1980). Importantly, Figure 2 also reveals an effect of spoken word properties, posttarget reading being slower in the similar condition than in the identical and dissimilar conditions. Statistically, this was expressed in an interaction of critical position and spoken word type [$F(8,232) = 2.38, p < .025$].

Two paired comparisons were applied to each one of the five critical words to determine whether similar spoken words interfered more with target and posttarget reading than did dissimilar and identical words. The results revealed no reliable differences for the pretarget ($N - 1$) and the target (N) word (all $ps > .1$), but reliably longer gazes in the similar condition than in either the dissimilar or the identical condition for word $N + 1$ [$t(29) = 3.23$ and 3.74 , respectively; both $ps < .01$]. The differences between the similar condition and either the identical or the dissimilar condition were not reliable for words $N + 2$ and $N + 3$ (all $ps > .1$).

Discussion

The results reveal an immediate response of the oculomotor system to the presentation of an acoustic signal. Moreover, there was selective interference for similar spoken words during posttarget reading, presumably because visual and spoken words activated identical or structurally similar phonological representations. The technique thus appears to be a viable tool for the study of a visual word's phonological code.

Deleterious effects of sound similarity were confined, however, to the $N + 1$ word. Since a spoken companion word's articulation duration was generally longer than target viewing durations, information necessary for spoken word identification may have been provided while $N + 1$ was viewed. The confinement of similarity effects to $N + 1$ reading could thus indicate that effects of the target's phonological representation emerged only when the spoken word was recognized and that the contingent speech technique is insensitive to the time course of phonological code usage. Alternatively, a lack of experimental power may have prevented a more robust expression of spoken word effects during and after target reading. Information conveyed by an irrelevant spoken word may have been quickly rejected because it was delivered over speakers.

Speakers have a distinct spatial location that differs from the location of the visual target, and the acoustic signal can be degraded by environmental impurities, such as noise generated by the participant and by experimental hardware.

EXPERIMENT 2

Experiment 2 was conducted to determine whether the effects of a visual target's sound code emerged only after the spoken word had been fully recognized. The experiment was identical to Experiment 1, with two exceptions: Spoken word durations were decreased so that full articulation of the companion word was generally accomplished while the visual target was fixated, and all the spoken words were administered via headphones. If the effects of a spoken word on sentence reading occurred when it had been recognized, deleterious effects of a phonologically similar spoken word should occur during target viewing, rather than $N + 1$ viewing, in Experiment 2. Moreover, post-target words should show no or negligible effects of sound

similarity. Conversely, if lack of experimental power undermined the full expression of phonological similarity effects, target, $N + 1$, and even $N + 2$ reading may have been influenced by properties of the spoken word.

Method

Participants. Thirty-six undergraduate or graduate students with no history of reading disabilities participated in the study for course credit or payment. None of the students had participated in Experiment 1.

Materials, Design, and Data Analyses. These materials, design, and data analyses in Experiment 2 were identical to those in Experiment 1, except that the duration of spoken companion words was shortened. The compress routine of the speech-editing program was used to reduce articulation duration by one third. The mean of the reduced articulation duration (approximately 365 msec) was shorter than the mean of target gazes in Experiment 1 (approximately 440 msec). Two native speakers verified that all shortened speech tokens could be readily identified.

Apparatus. The same apparatus was used as that in Experiment 1, except that spoken words were presented via a pair of high-quality headphones (Sennheiser HD 525). Although this increased the amount of head gear, none of the readers reported being inconvenienced by it.

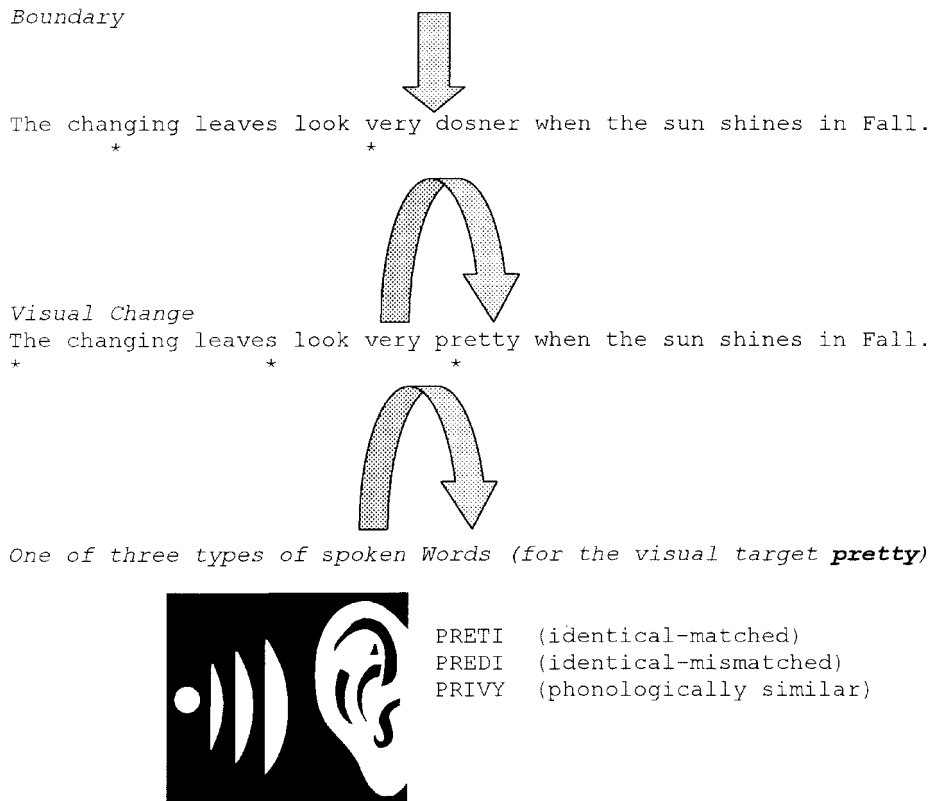


Figure 3. A sample sentence used in the second experiment, along with pretarget and target fixations, indicated by asterisks. The arrow shown for the top line reveals the sentence boundary. The movement of the eyes across the boundary, shown in the second line, revealed the visual target word. In addition, as is shown in the third line, it also led to the presentation of a spoken word that corresponded to the phonological transcript version of the visual target, a licensed articulation of the target that did not match its phonological transcription, or a phonologically similar word.

Results

Figure 2B shows $N - 1$ to $N + 3$ gazes as a function of the spoken companion word.

As in Experiment 1, presentation of the spoken word selectively increased target gazes, expressed in a robust position effect [$F(4,140) = 39.80, p < .01$], and spoken effects were a function of critical word location, expressed in a reliable interaction of spoken word type and visual word position [$F(8,280) = 4.04, p < .01$]. Effect sizes were larger, however, and more clear-cut. This was confirmed by paired comparisons that revealed longer gazes in the similar condition than in the identical condition for words N to $N + 2$ [$t(35) = 2.05, p < .05, t(35) = 3.41, p < .01$, and $t(35) = 3.01, p < .01$, respectively]. Gazes were also longer in the similar than the dissimilar condition while words $N + 1$ and $N + 2$ were viewed [$t(35) = 5.78, p < .01$, and $t(35) = 2.68, p < .025$, respectively]. No reliable differences emerged during pretarget and $N + 3$ viewing (all $ps > .1$).

Discussion

The results of Experiment 2 revealed immediate effects of spoken word type on target viewing and selective interference from phonologically similar spoken words during $N + 1$ and $N + 2$ reading. Phonological representations thus remained active after visual and spoken words had been recognized, presumably because they were used in working memory. These results indicate that the contingent speech technique has considerable promise for the specification of the time course of phonological information usage.

EXPERIMENT 3

Together, Experiments 1 and 2 indicated that the phonological representations of visual and spoken words are structurally similar and that they remain activated after visual word recognition has been achieved. In Experiment 3, the contingent speech technique was used to determine the nature of a visual target's phonological code—that is, whether it is akin to a phonological transcription of the graphemic code, as is generally assumed, or whether it has speech-like properties. For this, we paired each visual target word with one of two legal target articulations, one that matched and one that mismatched its grapheme-to-phoneme transcription. In some dialects of English, a medial /t/ in a word such as PRETTY is a hyperarticulated /t/. A flapped /d/-like production predominated in American English, however (Patterson & Connine, 2001), so that the word's articulation and grapheme-to-phoneme transcription mismatch. If activation of a visual word's phonological representation involves grapheme-to-phoneme conversion or the retrieval of a corresponding sound code, a matching spoken word with a fully articulated medial /t/ should be less disruptive and should be disruptive for a shorter duration than a mismatching spoken word with a corresponding medial flap.

In another experimental condition, viewing of a visual target was accompanied by the presentation of a phonologically similar word, and a dissimilar condition was created in which a set of target-like (filler) words was paired with presentations of dissimilar companion words. As in Experiment 2, similar spoken words were expected to increase target and subsequent posttarget viewing durations, relative to the matching (identical) condition, and the effects of dissimilar spoken words on filler trials were expected to dovetail effects of the matching (identical) condition.

Method

Participants. Forty-five undergraduate and graduate students from Binghamton University participated in the experiment either for course credit toward the fulfillment of a course requirement or for a monetary supplement. All the participants were native speakers of English, with normal hearing and normal or corrected-to-normal vision.

Materials. Visual target words consisted of 66 four- to seven-letter two-syllable words. All the targets were content words, with a mean CELEX frequency of 7 per million (Baayen, Piepenbrock, & van Rijn, 1993). The orthographic version of all the target words contained one or two near medial /t/s—for example, *pretty*, which during articulation in American English could be articulated either as a matched, unvoiced /t/, as in PRETI, or as a mismatched, voiced, d-like phoneme, as in PREDI (a *flapped* production). An additional set of 22 target-like bisyllabic filler words was selected, with a mean frequency of 13 per million. These filler words contained near medial /t/s—for example, *gently*—that are articulated as voiceless segments, as in GENTLY. Each target and filler word was embedded in sentence context, where it occurred near the sentence center. The sentences in Experiment 3 differed from the sentences in Experiments 1 and 2 in that they contained, on average, two additional post-target words.

Three spoken companion words were created for each visual target. This included two identical conditions, consisting of a matched version in which the medial letter t was articulated as a voiceless /t/, and a mismatched version, in which the medial t was flapped. The mean duration of matched and mismatched identical spoken words was 551 and 528 msec, respectively. The third spoken companion word different from the visual target in meaning but was phonologically similar to it—for example, the target *pretty* was paired with PRIVY. Similar spoken words and the two spoken versions of target words were matched on a wide range of properties, including word frequency (6 per million), syllable structure, and word type. Their mean articulation duration was 607 msec. A sample trial for the three experimental conditions is shown in Figure 3. All the visual filler words were paired with a bisyllabic dissimilar companion word. All the dissimilar companion words were content words with a mean frequency of 21 per million; their mean articulation duration was 672 msec.

Procedure, Design, and Data Analyses. The procedure, design, and data analyses in Experiment 3 were identical to those in Experiment 2, except that the spoken words were articulated at a normal tempo. Three lists were created so that each target word was presented with a different spoken companion word (identical-matched, identical-mismatched, or similar) on each list. Filler words were always accompanied by dissimilar spoken words, and filler trials are also referred to as dissimilar trials. Filler trials were included so that the viewing of a word with a near medial /t/ would not always be combined with an identical or similar-sounding spoken companion word and to provide means of comparing the results of Experiment 3 with the results of Experiments 1 and 2.

An omnibus analysis of variance (ANOVA) was used to examine spoken word effects for experimental and filler words. This was fol-

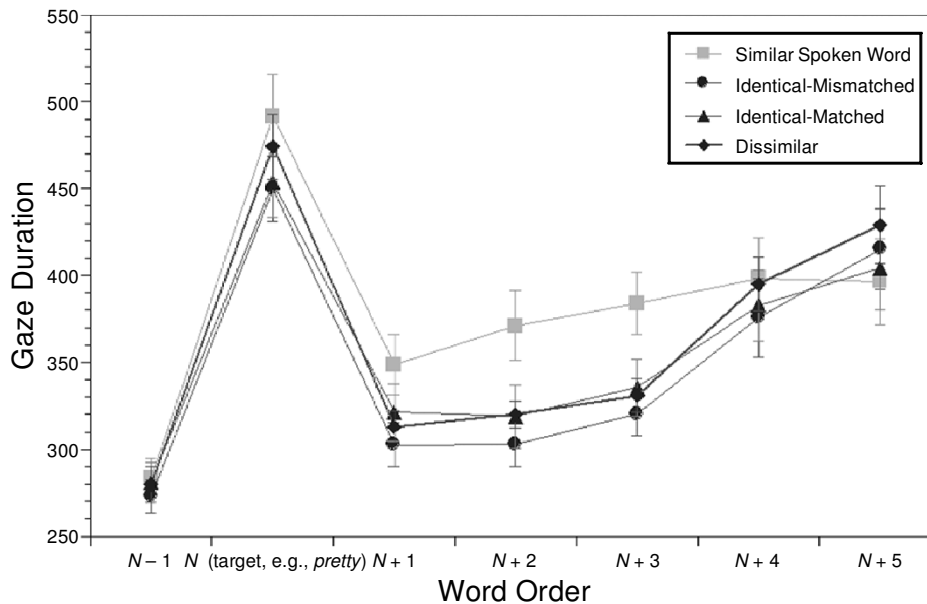


Figure 4. Pretarget, target, and posttarget gaze durations as a function of spoken word type. The fifth posttarget word generally coincided with the sentence ending.

lowed by more specific analyses that involved contrasts of the two identical conditions, to determine whether the identical-matched condition was less disruptive than the identical-mismatched condition, and by contrasts of the two identical conditions with the similar conditions. A final supplementary set of analyses contrasted the two identical conditions with the dissimilar (filler) condition. This involved the comparison of spoken word effects for different but well-matched sets of visual words.

Results

The time spent viewing pretarget, target, and the next five posttarget words, is shown in Figure 4 as a function of spoken word type. The figure also shows the corresponding viewing durations for dissimilar filler trials.

As in Experiments 1 and 2, presentation of a spoken word resulted in an immediate increase in gaze durations, and the type of spoken word determined the time course of spoken word effects. This was expressed in a reliable effect of sentence location [$F(6,264) = 45.71$] and an interaction of location and spoken word type [$F(18,792) = 3.00$, both $ps < .01$] in the omnibus ANOVA. Supplementary analyses, for each of the seven sentence locations, revealed a reliable spoken word effect for the target [$F(3,132) = 12.69$], word $N + 1$ [$F(3,132) = 9.07$], word $N + 2$ [$F(3,132) = 9.36$], and word $N + 3$ [$F(3,132) = 4.58$, all $ps < .01$]. Spoken word effects were not reliable for the pretarget word, word $N + 4$, and word $N + 5$ (all $ps > .1$). As in Experiment 2, spoken word effects emerged immediately, during target viewing, and continued after the visual and spoken words had been recognized.

To determine whether an identical-matched spoken word was less disruptive than an identical-mismatched

word, gazes for words N (target) to $N + 3$ were examined as a function of spoken word type (matched vs. mismatched). The results were negative. Instead of providing an advantage, matched spoken words yielded slightly longer viewing durations, although the difference did not approach significance ($F_s < 1$).

The contrasts of the identical-matched and identical-mismatched conditions with the similar condition for the words N to $N + 3$ revealed a main effect of spoken word type, gazes being longer in the similar condition than in either the identical-matched or the identical-mismatched condition [$F(1,44) = 28.67$ and 50.71 , respectively]. The interaction of word location and spoken word type was negligible in both analyses. As in Experiment 2, phonological similarity interfered with target and posttarget reading.

The final supplementary set of contrasts compared spoken word effects in the matched and mismatched conditions with spoken word effects in the dissimilar condition (filler trials). It revealed significant interactions of sentence location and spoken word type in both analyses [$F(3,132) = 5.21$ and 5.62 , respectively; both $ps < .01$]. As in Experiment 2, dissimilar spoken words yielded relatively large immediate effects (during target and control word viewing) and relatively small delayed posttarget effects.

Discussion

Experiment 3 replicated key aspects of Experiment 2. Again, phonological similarity interfered with target and posttarget reading. Deleterious effects of similarity were

longer lasting, now including word $N + 3$, presumably because of the larger number of posttarget words in Experiment 3 than in Experiment 2. This suggests that the phonological code of a visual word may remain active until a sentence boundary is approached—that is, until wrap-up effects take over.

The results also showed that a spoken companion word that mismatched the grapheme-to-phoneme transcription of the visual target was no more disruptive than a spoken companion word that contained a matching phoneme. This may occur because the graphemic form of a visual word activates multiple phonological forms of a visual target, including a form that has speech-like properties.

GENERAL DISCUSSION

The main goal of the present study was to examine the promise of a novel eye-movement-contingent speech technique that offers the experimenter precise control over the availability of linguistic information under relatively natural task conditions. In contrast to other extant eye-movement-contingent techniques, successful implementation of the experimental manipulation does not depend on the duration of a target's fixation, and effects are not compromised by potential effects of visual display changes during a fixation on eye movements. In addition, the contingent speech technique can be used to implement linguistic manipulations that are difficult, if not impossible, to implement via other experimental methods available in psycholinguistic research.

Experiments 1–3 used the technique to determine the life cycle and nature of phonological representations during reading. One major finding showed that spoken word type effects extended substantially beyond the recognition of the visual target and its spoken companion word, similar spoken words being more disruptive than identical or phonologically dissimilar spoken words, thus indicating that the method can be used to examine recognition and working memory processes. The corresponding effect sizes were relatively large, often exceeding 50 msec. The second major finding showed that a spoken variant of the visual target that matched its grapheme-to-phoneme transcription was no less disruptive to target and posttarget reading than was a mismatching spoken variant of the target, suggesting that readers activate multiple phonological forms of a visual word, including a form that has speech-like properties.

Other properties of the spoken companion word can be manipulated, and the pattern of interference caused by different types of irrelevant spoken companion words can be used to identify functional representations and the time course of their activation. Preliminary data from our laboratory show, for instance, longer target and posttarget viewing durations when the meanings of the target and its spoken companion word are similar than when they are dissimilar. Preliminary results from another study, in which we manipulated the proximity of a target word to a

sentence boundary, suggest that phonological codes remain active until the boundary is reached.

In principle, the contingent speech technique may also be used in areas other than reading—for example, in studies of picture processing or picture speech interaction, where the viewing of selected picture content could be used to initiate the controlled presentation of a selected acoustic signal. In applied settings, the procedure may be used to investigate the effects of different types of noise or acoustic information on visual information processing and/or decision making. The technique could thus provide a powerful addition to the arsenal of methods in current eye movement research.

REFERENCES

- BAAYEN, R. H., PIEPENBROCK, R., & VAN RIJN, H. (1993). *The Celex lexical database*. (CD-ROM). University of Pennsylvania, Philadelphia: Linguistic Data Consortium.
- BIRCH, S., POLLATSEK, A., & KINGSTON, J. (1998). The nature of the sound codes accessed by visual language. *Journal of Memory & Language*, **38**, 70-93.
- CRAEMER, A., & RADACH, R. (2002). *Software for the control of eye movement experiments using EyeLink eye tracking systems* (Technical Report). Aachen: Technical University of Aachen.
- FROST, R. (1998). Toward a strong phonological theory of visual word recognition: True issues and false trials. *Psychological Bulletin*, **123**, 71-99.
- HOCHBERG, J. (1976). Toward a speech-plan eye-movement model of reading. In R. A. Monty & J. W. Senders (Eds.), *Eye movements and psychological processes* (pp. 374-396). Hillsdale, NJ: Erlbaum.
- INHOFF, A. W., & RADACH, R. (1998). Definition and computation of oculomotor measures in the study of cognitive processes. In G. Underwood (Ed.), *Eye guidance in reading and scene perception* (pp. 29-54). Amsterdam: Elsevier.
- INHOFF, A. W., RADACH, R., EITER, B., & JUHASZ, B. (in press). Parafoveal processing: Distinct subsystems for spatial and linguistic information. *Quarterly Journal of Experimental Psychology*.
- INHOFF, A. W., STARR, M., LIU, W., & WANG, J. (1998). Eye-movement-contingent display changes are not compromised by flicker and phosphor persistence. *Psychonomic Bulletin & Review*, **5**, 101-106.
- INHOFF, A. W., & WEGER, U. (in press). Advancing the methodological middle ground. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movements*. Amsterdam: Elsevier.
- JUST, M. A., & CARPENTER, P. A. (1980). A theory of reading: From eye fixations to comprehension. *Psychological Review*, **87**, 329-354.
- KENNEDY, A., RADACH, R., HELLER, D., & PYNTE, J. (Eds.) (2000). *Reading as a perceptual process*. Amsterdam: Elsevier.
- MCCONKIE, G. W., & RAYNER, K. (1975). The span of the effective stimulus during a fixation in reading. *Perception & Psychophysics*, **17**, 578-586.
- MCCONKIE, G. W., WOLVERTON, G. S., & ZOLA, D. (1984). Instrumentation considerations in research involving eye movement contingent stimulus control. In A. G. Gale & F. Johnson (Eds.), *Theoretical and applied aspects of eye movement research* (pp. 39-47). Amsterdam: Elsevier.
- PATTERSON, D., & CONNINE, C. M. (2001). A corpus analysis of variant frequency in American English flap production. *Phonetica*, **58**, 254-275.
- RADACH, R., & KENNEDY, A. (in press). Theoretical perspectives on eye movements in reading: Past controversies, current deficits and an agenda for future research. *European Journal of Cognitive Psychology*.
- RAYNER, K. (1975). The perceptual span and peripheral cues in reading. *Cognitive Psychology*, **7**, 65-81.
- RAYNER, K. (1998). Eye movements in reading and information

- processing: 20 years of research. *Psychological Bulletin*, **124**, 372-422.
- REINGOLD, E. M., & STAMPE, D. M. (2000). Saccadic inhibition and gaze contingent research paradigms. In A. Kennedy, R. Radach, D. Heller, & J. Pynte (Eds.), *Reading as a perceptual process* (p. 119-145). Amsterdam: Elsevier.
- SERENO, S. C., & RAYNER, K. (1992). Fast priming during eye fixations in reading. *Journal of Experimental Psychology: Human Perception & Performance*, **18**, 173-184.
- STARR, M., & FLEMING, K. (2001). A rose by any other name is not the same: The role of orthographic knowledge in homophone confusion errors. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, **27**, 744-60.

(Manuscript received March 20, 2002;
revision accepted for publication August 6, 2002.)