Available online at www.sciencedirect.com
ScienceDirect
Journal of Memory and Language 57 (2007) 232–251
Journal of Memory and Language
www.elsevier.com/locate/jml

Attention and gaze control in picture naming, word reading, and word categorizing ☆

Ardi Roelofs *

Nijmegen Institute for Cognition and Information and F.C. Donders Centre for Cognitive Neuroimaging, Radboud University Nijmegen, Spinoza Building B.02.32, Montessorilaan 3, 6525 HR Nijmegen, The Netherlands

Received 12 June 2006; revision received 10 October 2006
Available online 1 December 2006

Abstract

The trigger for shifting gaze between stimuli requiring vocal and manual responses was examined. Participants were presented with picture–word stimuli and left- or right-pointing arrows. They vocally named the picture (Experiment 1), read the word (Experiment 2), or categorized the word (Experiment 3) and shifted their gaze to the arrow to manually indicate its direction. The experiments showed that the temporal coordination of vocal responding and gaze shifting depends on the vocal task and, to a lesser extent, on the type of relationship between picture and word. There was a close temporal link between gaze shifting and manual responding, suggesting that the gaze shifts indexed shifts of attention between the vocal and manual tasks. Computer simulations showed that a simple extension of WEAVER++ [Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition*, *42*, 107–142.; Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review*, *110*, 88–125.] with assumptions about attentional control in the coordination of vocal responding, gaze shifting, and manual responding quantitatively accounts for the key findings.

Keywords: Attention; Computational modeling; Executive control; Eye tracking; Speech production

Individuals often perform two tasks concurrently or in close succession, such as talking while driving a car, preparing a meal, or manipulating a computer mouse. The ability to cope with such dual-task situations depends on an individual's ability to coordinate cognitive processes across the tasks at hand (e.g., Allport, 1980a, 1980b, 1987; Meyer & Kieras, 1997a, 1997b; Monsell, 1996; Pashler, 1994; Shallice, 1988; Telford, 1931; Welford, 1952), which is often called attentional or executive control. The problem of attentional control concerns "the computational mechanisms by which attentional engagement is established, coordinated, maintained, interrupted, and redirected, both in spatial and nonspatial terms, in the preparation and control of action" (Allport, 1989, pp. 662–663). An important attentional control process involves "the direction of gaze" (Monsell, 1996, p. 96) or "ocular orientation" (Meyer & Kieras, 1997a, p. 26). Eye movements need to be coordinated between tasks if the stimuli for the

☆ I am indebted to Bicoor Bolla-Bong for his help in running the experiments, and to Martijn Lamers, Stephen Lupker, Antje Meyer, Keith Rayner, and anonymous reviewers for helpful comments. The preparation of the article was supported by a VICI grant from the Netherlands Organisation for Scientific Research (NWO).
* Fax: +31 24 3616066.
E-mail address: A.Roelofs@nici.ru.nl

0749-596X/$ - see front matter © 2006 Elsevier Inc. All rights reserved.
doi:10.1016/j.jml.2006.10.001

two tasks have different spatial positions. The present article addresses the question how eye movements are coordinated between vocal and manual tasks involving spatially separated stimuli.

In the second half of the nineteenth century, well before the modern era of cognitive and brain sciences, two pioneers of speech and language production research, Donders and Wundt, studied eye movements (Donders, 1870; Wundt, 1862). Donders was the first person to measure vocal naming latencies, and he developed a model for eye movements (i.e., a mechanical model demonstrating what came to be called ‘‘Donders’ Law’’). Whereas before those days the eyes used to be poetically called a window to the soul, Wundt (1897) took gazes to be a window into the operation of the attention system. As Wundt (1897) reasoned in his *Outlines of psychology*, visual acuity is best at the center of eye fixation (by 5° from the center, acuity has diminished about 50%). Therefore, to bring aspects of the visual world in the focus of attention, eye fixations are directed to those visual aspects that are of interest. This makes a shift of gaze between two visual stimuli an overt sign of the orienting of attention (cf. Hoffman & Subramaniam, 1995; Kowler, Anderson, Dosher, & Blaser, 1995; Shepherd, Findlay, & Hockey, 1986), although attention and eye movements can sometimes be dissociated in simple signal detection and identification tasks (e.g., Posner, 1978, 1980). See Rayner (1998) for a review of the literature.

A prominent model of attentional control processes in dual-task performance developed by Meyer and Kieras (1997a, 1997b) assumes that a saccadic eye movement from one task stimulus (Task 1) to another (Task 2) is issued directly by the attentional control process. The model is called the strategic response-deferment (SRD) model and it has been implemented in the executive-process interactive control (EPIC) architecture (Meyer & Kieras, 1997a). In particular, Meyer and Kieras (1997a, 1997b) supposed that when the perceptual processing of a visual stimulus for the first task has progressed far enough, the eyes are instructed to move to the location of the visual stimulus for the second task. The assumption about the early, perception-based move of gaze was used in tests of EPIC-SRD through computer simulations of existing experimental findings.

For example, in one of the simulated experiments, participants had to give a vocal response to one of two letters presented on the left side of a computer screen (Task 1) and they had to give a manual response to one of two digits presented on the right side of the screen (Task 2). The vocal and manual response latencies were measured. In computer simulations, EPIC’s attentional control process requested a saccadic eye movement from the Task 1 stimulus to the Task 2 stimulus immediately after the onset of the Task 1 stimulus. Thus, the shift of gaze was requested before selecting the response for Task 1. This assumption about gaze shifts proved to be useful in fitting EPIC-SRD quantitatively to the empirical data about the vocal and manual response latencies. However, although Meyer and Kieras (1997a, 1997b) used this assumption about gaze shifts in their simulations, it was not explicitly tested. Information about eye movements was not available from the simulated empirical studies.

Sanders (1998; Sanders & Lamers, 2002; Van Duren & Sanders, 1995) made an assumption about the control of eye movements that was similar to that implemented in EPIC. According to Sanders (1998), gaze shifts are based on perceptual processing and are initiated before response selection. Empirical evidence was provided for this claim. For example, in one of the experiments reported by Van Duren and Sanders (1995), participants saw one of the digits 1, 2, 3, or 4 on the left side of the screen. They had to respond by pressing one of four keys. The order in which the digits were assigned to response keys was from left to right, a compatible stimulus–response mapping, or from right to left, an incompatible stimulus–response mapping. Evidence suggests that stimulus–response compatibility affects response selection (e.g., Kornblum, Hasbroucq, & Osman, 1990; see Sanders, 1998, for a review of the literature). A go/no-go stimulus presented on the right side of the screen indicated whether the response had to be executed or not. Van Duren and Sanders (1995) observed that the gaze durations to the digits were unaffected by stimulus–response compatibility. This finding suggests that response selection occurs during the saccade from the digit to the go/no-go symbol. Similarly, Sanders and Lamers (2002) observed that the effect of incongruent versus congruent flanker letters in the classic Eriksen task was not reflected in the latency of gaze shifts to a go/no-go symbol. According to Sanders (1998; Sanders & Lamers, 2002; Van Duren & Sanders, 1995), a gaze shift is determined by perceptual factors and occurs before response selection (the presumed functional locus of the flanker effect).

The assumption made by Meyer and Kieras (1997a, 1997b) and Sanders (1998) about saccadic eye movements does not accord with recent findings on vocal responding, which suggest that gaze shifts depend on the time to plan spoken words (e.g., Griffin, 2001; Meyer, Sleiderink, & Levelt, 1998). In general, when individuals vocally refer to a number of seen objects, gaze durations appear to be tightly linked to the timing of linguistic planning processes for each object (e.g., Bock, Irwin, Davidson, & Levelt, 2003; Griffin & Bock, 2000; Meyer et al., 1998). For example, when speakers are asked to name two objects in a row, they look longer at first-to-be-named objects with two- than with one-syllable names even when the object recognition times do not differ (Meyer, Roelofs, & Levelt, 2003). Zelinsky and Murphy (2000) made a similar observation for a task that did not demand overt naming but that encouraged verbal encoding. The WEAVER++ model of vocal response planning (e.g., Levelt, Roelofs, & Meyer, 1999; Roelofs, 1992, 1997, 2003) divides the planning process into lemma retrieval (response selection) and word-form encoding (response programming), with word-form encoding further divided into morphological encoding, phonological encoding, and phonetic encoding. The effect of phonological length suggests that the shift of gaze from one object to another is initiated only after the phonological form of the object name has been encoded.

Experimental manipulations may have different effects on vocal response latencies and gaze shifts. Levelt and Meyer (2000) and Korvorst, Roelofs, and Levelt (2006) observed that gaze durations may reflect the phonological length of the utterance even when vocal response latencies do not. Levelt and Meyer (2000) instructed their participants to describe colored left and right objects (e.g., a big red scooter and a ball) in a simple or in a complex way. Participants either had to respond with "the scooter and the ball" or "the *big red* scooter and the ball." The gaze durations for the left object (the scooter) were much shorter for the simple utterances than for the complex utterances. However, the vocal response latencies did not differ between the two utterance types. Furthermore, the shift of gaze to the right object was initiated before articulation onset for the simple utterances, but after articulation onset for the complex utterances. This suggests that the shift of gaze, but not the onset of articulation, is triggered by the completion of phonological encoding of the utterance referring to the first object.

Why do gaze shifts seem to be initiated before response selection in the task situations of Meyer and Kieras (1997a, 1997b) and Sanders (1998), but during response programming in language production tasks (e.g., Bock et al., 2003; Griffin & Bock, 2000; Meyer et al., 1998, 2003)? It is possible that the difference in experimental results is related to two important concerns in the management of plans of action: "minimizing resource consumption" and "avoiding plan failure" (Charniak & McDermott, 1985, p. 488).

As concerns the minimization of resource consumption, a major difference between studies is whether or not Tasks 1 and 2 used the same effector system, leading to differences in the need for response buffering. In the experiment simulated by Meyer and Kieras (1997a), the Task 1 response was vocal and the Task 2 response was manual. In the experiments of Van Duren and Sanders (1995) and Sanders and Lamers (2002), participants had to give only a single response based on both the left and right stimuli. In contrast, in the experiments of Meyer et al. (2003) and Korvorst et al. (2006), participants had to vocally name two visual stimuli. Articulation is a slow process. Pronouncing a word can easily take half a second or more (cf. Levelt & Meyer, 2000). Consequently, Task 2 planning may be completed well before articulation of the Task 1 response has been finished. This means that the vocal response for Task 2 needs to be buffered for a relatively long time. By adopting a serial strategy (i.e., only starting perception of the Task 2 stimulus when Task 1 planning is done), the use of buffering resources can be limited (Levelt & Meyer, 2000). A serial planning strategy is sometimes called "lockout scheduling" (Meyer & Kieras, 1997a, p. 20).

As concerns the avoidance of plan failure, "one way a plan can fail is for another plan to interfere with it" (Charniak & McDermott, 1985, p. 488). In the case of two vocal responses, planning the response for Task 2 may interfere with planning the response for Task 1. Shifting gaze away from an object while planning its name may lead to interference from seeing other objects whose names may inadvertently be activated (Meyer & Van der Meulen, 2000; Zelinsky & Murphy, 2000). According to Zelinsky and Murphy (2000), "the tight visual–verbal coupling is likely designed to minimize interference" (p. 130). The interference hypothesis is supported by evidence suggesting that context pictures activate their names (e.g., Morsella & Miozzo, 2002; Navarrete & Costa, 2005; Roelofs, 2006). Moreover, it is supported by so-called environmental intrusion errors, whereby irrelevant information about what speakers are looking at is inadvertently included in an utterance (Harley, 1984). With vocal Task 1 responses and manual Task 2 responses (Meyer & Kieras, 1997a, 1997b), a shift of gaze after the perception of the Task 1 stimulus should not lead to much interference from Task 2 response planning on Task 1 response planning, whereas there may be much more interference when both responses are vocal. Thus, by adopting a serial strategy, the planning of the first vocal response is protected against interference from planning the second vocal response.

A final distinction between the experimental situations of Meyer and Kieras (1997a, 1997b) and Sanders (1998) and those of the language production research (e.g., Bock et al., 2003; Meyer et al., 1998, 2003; Griffin & Bock, 2000) is a difference in task demands. In the experiment modeled by Meyer and Kieras (1997a, 1997b), participants had to give a vocal response to one of two letters presented on the screen, whereas in the experiments of Meyer et al. (2003), 32 different pictured objects had to be named. Evidence suggests that the number of responses affects the difficulty of response selection (e.g., Kornblum et al., 1990; see Sanders, 1998, for a review of the literature). For example, Sternberg (1969) found that effects of the number of responses are additive with stimulus legibility and interact with stimulus–response compatibility. Moreover, with several possible responses, participants might engage in more extensive self-monitoring of response planning than when there are only a few responses. If shifts of gaze reflect the task demands, earlier shifts of gaze would be expected when response selection is easy and attention demands are low (e.g., Meyer & Kieras, 1997a, 1997b) compared to when response selection is harder and attention demands are higher (e.g., Meyer et al., 2003).

The present experiments

The aim of the experiments reported in this article was to examine what the trigger is for moving the eyes between visual stimuli in task situations involving vocal and manual responding. As in the experiment simulated by Meyer and Kieras (1997a), Task 1 required a vocal response and Task 2 required a manual response. A Task 2 with manual responses was chosen in order to avoid the need for vocal response buffering and to minimize verbal interference between the planning of Tasks 1 and 2 responses. The experiments were designed to examine the question whether saccadic eye movements are triggered before response selection (Meyer & Kieras, 1997a, 1997b; Sanders, 1998) or during vocal response programming (e.g., Meyer et al., 2003). Moreover, the role of task demands was addressed. In order to be able to assess when saccadic eye movements are triggered, the difficulty of vocal response selection was manipulated. This was done by having participants vocally respond to picture–word combinations. Evidence suggests that picture–word interference effects reflect the difficulty of response selection in vocal response planning (for reviews, e.g., MacLeod, 1991; Roelofs, 2003). Participants were presented with picture–word stimuli displayed on the left side of a computer screen and left- or right-pointing arrows displayed on the right side of the screen. The arrows < and > were flanked by two Xs on each side. Fig. 1 illustrates the visual displays used in the experiments. The picture–word stimulus and the arrow were presented simultaneously on the screen. The participants' tasks were to vocally respond to the picture–word stimulus (Task 1) and to manually indicate the direction in which the arrow was pointing by pressing a left or right button (Task 2). Eye movements were recorded in order to determine the onset of the shift of gaze between the picture–word stimulus and the arrow.

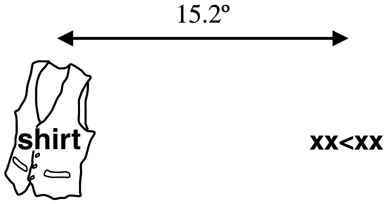

Fig. 1. Illustration of the stimulus displays used in the experiments, approximately on scale.

In Experiment 1, participants named the pictures of picture–word stimuli while trying to ignore the written words superimposed onto the pictures. The picture–word stimuli had varying types of relatedness between picture and word. For example, participants said "vest" in response to a pictured vest, while trying to ignore the word SHIRT (the semantic condition), the word CASTLE (the unrelated condition), the word VEST (the identical condition), or a series of Xs (the control condition). In Experiment 2, participants read aloud the words of the picture–word stimuli while trying to ignore the pictures. In Experiment 3, participants generated a category name in response to the words (e.g., they said "clothing" in response to the word VEST) while trying to ignore the pictures. The number of responses in Experiments 1 and 2 was exactly the same (i.e., 32), but the stimulus–response compatibility differed between experiments. The compatibility was lower for picture naming in Experiment 1 than for word reading in Experiment 2, because the orthography of a word provides information about its pronunciation, whereas the form of a pictured object does not (e.g., Kornblum et al., 1990). Furthermore, the number of responses was larger in picture naming in Experiment 1 than in word categorizing in Experiment 3, namely 32 vs. 8, whereas in both experiments the stimulus–response compatibility was lower than in word reading in Experiment 2. Moreover, distractor effects are different between tasks.

Previous research (e.g., Glaser & Düngelhoff, 1984; Glaser & Glaser, 1989; Lupker, 1979; Rayner & Springer, 1986; Smith & Magee, 1980) showed that participants are slower in naming a picture with a semantically related word superimposed (e.g., saying "vest" to a pictured vest with the written word SHIRT superimposed) than in naming the picture with a series of Xs superimposed in the control condition. Furthermore, participants are often faster than control when picture and word agree in the identical condition (e.g., saying "vest" to a pictured vest with the identical word VEST superimposed). Moreover, a semantic effect is obtained. Participants are slower in naming a picture (e.g., saying "vest" to a pictured vest) with a semantically related word superimposed (e.g., SHIRT) than with an unrelated word (e.g., CASTLE). When the task is to read aloud the words and to ignore the pictures, there is no interference from semantically related pictures or facilitation from identical pictures relative to control. There is also no semantic effect, that is, there is no difference in effect between semantically related and unrelated picture distractors (e.g., Glaser & Düngelhoff, 1984). However, when the words have to be categorized (e.g., saying "clothing" to the word VEST), the response latencies with picture distractors are shorter in the semantic than in the unrelated condition, whereas the latencies in the semantic, identical, and control conditions do not differ (e.g., Glaser & Düngelhoff, 1984). Computational accounts of the distractor effects are provided by WEAVER++ (e.g., Levelt et al., 1999; Roelofs, 1992, 2003).

If gaze shifts are triggered before response selection (Meyer & Kieras, 1997a, 1997b), the vocal response latencies but not the gaze shift latencies should reflect the picture–word interference effects. Such a finding would be in agreement with the assumption made by Meyer and Kieras (1997a, 1997b) and the empirical findings of Van Duren and Sanders (1995) and Sanders and Lamers (2002). Alternatively, gaze shifts may be triggered upon completion of phonological encoding of the vocal response (e.g., Meyer et al., 2003). If so, the distractor effects on the gaze shifts should pattern with those on the vocal responses. It may be argued that Levelt and Meyer (2000) and Korvorst et al. (2006) found that effects of utterance length are sometimes obtained in the latencies of gaze shifts even when they are not obtained in the vocal response latencies. However, in the present experiments, all responses were tested in all distractor conditions. Thus, differences in distractor effects on the vocal responses and gaze shifts cannot be due to differences in target-word length between distractor conditions. Thus, if the completion of phonological encoding for the vocal response triggers a gaze shift, the picture–word effects obtained for the vocal responses should be propagated into the gaze shifts. Such a finding would be in agreement with the data from earlier language production studies in which the two task stimuli both required a vocal response (e.g., Meyer et al., 2003). Moreover, such a finding would indicate that gaze shifts depend on vocal response programming even when the Task 2 response is a manual one. This would suggest that the avoidance of response buffering and the prevention of interference from the second response are not the only reasons for a late gaze shift. Rather, the completion of phonological encoding of the vocal response would appear to be a major determinant of gaze shifts.

A third, related possibility is that gaze shifts reflect shifts of attentional engagement from Tasks 1 to 2. In Experiments 1 and 2, participants produced exactly the same words, only the tasks differed between experiments. If gaze shifts are determined by the completion of a particular aspect of word planning (Levelt & Meyer, 2000), the coordination of vocal responding and gaze shifting should be the same in the experiments. However, if gaze shifting depends on the attention demands of a task, gaze shifts should occur earlier relative to articulation onset with word reading in Experiment 2 (low attention demands) than with picture naming in Experiment 1 (high attention demand). Moreover, relative to the onset of articulation, gaze shifts should occur earlier with word categorizing in Experiment 3 (few responses) than with picture naming in Experiment 1 (many responses).

If gaze shifts reflect the disengagement of attention from Task 1, there should be a tight link between the distractor effects on the gaze shifts and the manual responses. Such a link is not necessary, because it has been shown that attention and eye movements can be dissociated (e.g., Posner, 1980). While individuals cannot move their eyes to one location while paying full attention to another (i.e., shifts of eye position require shifts in attention), the focus of attention can shift without an eye movement (Hoffman & Subramaniam, 1995; Kowler et al., 1995; Shepherd et al., 1986). As Wundt argued and later studies confirmed, "this happens when we voluntarily concentrate our attention on a point in the eccentric regions of the field of vision" (Wundt, 1897, p. 213). Thus, it is possible that the eyes remain fixated on the left picture–word stimulus while attention is covertly moved to the right side of the screen before a gaze shift. Posner and colleagues (1980; Posner, Snyder, & Davidson, 1980) showed that a cue directing attention to a spatial position speeds up responding to a stimulus that subsequently appears in that location, even when the eyes remain fixated on the cue. Faster responses were obtained for simply detecting the appearance of the stimulus and for identifying it as a digit or letter. Tsal (1983) observed that the response gets faster as the time interval between cue and stimulus increases. Thus, if attention moves to the location of the arrow before the eyes do in the present Experiments 1–3, this should speed up the manual responses. This should hold even when peripheral vision of the arrow is insufficient to allow for visual pre-processing of the arrow. The effect of precueing is obtained regardless of whether the distance between cue and stimulus is 0.5° or 25° of visual angle (Posner, 1980; Posner et al., 1980). If an attention shift happened independently of a shift of gaze in the present Experiments 1–3, the gaze durations would not reflect the period of time that attention is given to the left picture–word stimulus. However, if the differences in gaze durations were preserved by the manual response latencies, this would be evidence that the shift of gaze corresponds to a shift of attention. The latter would be in line with a suggestion by Rayner (1998), who argued that "although we can easily decouple the locus of attention and eye location in simple discrimination tasks (Posner, 1980), in complex information processing tasks such as reading, the link between the two is probably quite tight" (p. 375).

Experiment 1

In the first experiment, the participants' tasks were to name the pictured item while ignoring the superimposed word and to manually indicate the direction in which the arrow was pointing. The latencies of the vocal responses, gaze shifts, and manual responses were recorded. The superimposed words were either semantically related to the picture (e.g., the word SHIRT printed on a picture of a vest), unrelated to the picture (e.g., CASTLE printed on a picture of a vest), identical to the picture (e.g., VEST), or a control string of Xs was superimposed (cf. Glaser & Düngelhoff, 1984). To demand that the arrows were fixated and to minimize the chance that participants could identify the direction of the arrows by their peripheral vision, the arrows were flanked by two Xs on each side, yielding XX < XX and XX > XX as stimuli.

Method

Participants

Each experiment was carried out with a different group of 16 paid participants from the pool of the Max Planck Institute for Psycholinguistics in Nijmegen. All participants were young adults, who were native speakers of Dutch. None of the participants took part in more than one experiment.

Materials and design

From the picture gallery available at the Max Planck Institute, 32 pictured objects from 8 different semantic categories were selected together with their basic-level names in Dutch. The Appendix A lists the materials. Furthermore, four additional pictures from two semantic categories (different from the eight experimental categories) were selected as practice items. The pictures were white line drawing on a black background. They were digitized and scaled to fit into a virtual frame of 10×10 cm. On average, the pictures subtended 8.7° horizontally and 8.7° vertically at a viewing distance of 66 cm. The distractor words were presented in 36-point lowercase Arial font. On average, the words subtended 1.3° vertically and 5.2° horizontally. The arrows, XX < XX and XX > XX, were presented in 28-point uppercase Arial font, subtending 0.9° vertically and 3.5° horizontally. The horizontal distance between the middle of the picture–word stimuli and the arrow stimuli was 15.2°.

The first independent variable was *measure*: vocal response, gaze shift, manual response. The second independent variable was *distractor*. Each target picture was combined with the corresponding word (the identical condition), with a word from the same semantic category (the semantic condition), with a word from another semantic category (the unrelated condition), or with a series of five Xs (the control condition) following Glaser and Düngelhoff (1984). These conditions were created by recombining pictures and words. All pictures and words occurred equally often in all conditions. The Appendix A gives the picture–word pairings. The word and the picture name always had the same grammatical gender. There were two items with neuter and two with nonneuter gender in each semantic category. A participant received 32 picture–word pairings in each of the four distractor conditions, yielding 128 picture–word stimuli in total. The order of presenting the stimuli across trials was random, except that repetitions of pictures and words on successive trials were not permitted.

Apparatus

Materials were presented on a 39 cm ViewSonic 17PS screen. Eye movements were measured using an SMI EyeLink-HiSpeed 2D headband-mounted eyetracking system (SensoMotoric Instruments GmbH, Teltow, Germany). The eyetracker was controlled by a Pentium 90 MHz computer. The experiment was run under the Nijmegen Experiment Setup (NESU) with a NESU button box on a Pentium 400 MHz computer. The participants' utterances were recorded over a Sennheiser ME400 microphone to a SONY DTC55 digital audio tape (DAT) recorder. Vocal response latencies were measured using an electronic voice key. Vocal and manual response latencies were measured from picture–word onset.

Procedure

The participants were tested individually. They were seated in front of the computer monitor, a panel with a left and a right push button, and the microphone. The distance between participant and screen was approximately 66 cm. Participants were given written instructions telling them how their eyes would be monitored and what the task was. The experimenter also orally described the eyetracking equipment and restated the instructions. The participants were told that they had to name the picture of picture–word stimuli presented on the left side of a computer screen and manually respond by pressing a left or right button in response to the arrows XX < XX or XX > XX presented on the right side of the screen. The instructions requested that the vocal responses should have earlier onsets than the manual responses (cf. Meyer & Kieras, 1997a, 1997b; Pashler, 1994). The participants were asked to respond as fast as possible without making mistakes. To familiarize them with the words and the pictures, the participants received a booklet showing them all pictures used in the experiment together with the expected names.

When a participant had read the instructions and had studied the picture booklet, the headband of the eyetracking system was placed on the participant's head and the system was calibrated and validated. For pupil-to-gaze calibration, a grid of three by three positions had been defined. During a calibration trial, a fixation target appeared once, in random order, in each of these positions for one second. Participants were asked to fixate upon each target until the next target appeared. After the calibration trial, the estimated positions of the participant's fixations and the distances from the fixation targets were displayed to the experimenter. Calibration was considered adequate if there was at least one fixation within 1.5° of each fixation target. When calibration was inadequate, the procedure was repeated, sometimes after adjusting the eye cameras. Successful calibration was followed by a pupil-to-gaze validation trial. For the participants, this trial did not differ from the calibration trial, but the data collected during the validation trial were used to estimate the participants' gaze positions, and the error (i.e., the distance between the estimated gaze position and the target position) was measured. Validation was considered completed if the average error was below 1.0° and the worst error below 1.5°. Depending on the result of the validation trial, the calibration and validation trials were repeated or testing began.

After successful calibration and validation, a block of 32 practice trials was administered. This was followed by the 128 experimental trials. The structure of a trial was as follows. A trial started by the simultaneous presentation of the left (picture–word) and right (arrow) stimuli. The stimuli remained on the screen until the participant pushed the button in response to the arrow. The latencies of the vocal and manual responses were measured from stimulus presentation onset. The picture, word, and arrow were presented in white on a black background. Before the start of the next trial there was a blank interval of 1.5 s. The position of the left and right eyes was determined every four milliseconds. Drift correction occurred automatically after every eight trials.

Analyses

To analyze the speakers' gaze shifts, their eye fixations were classified as falling within or on the outer contours of the left object or elsewhere. Although viewing was binocular and the positions of both eyes were tracked, only the position of the right eye was analyzed. The latency of a gaze shift was defined as the time interval between the beginning of the first fixation on the picture–word stimulus and the end of the last fixation before the first saccade was initiated to the arrow. Because the picture–word stimuli were always presented in the same position on the screen, there was no fixation point to indicate the position of the stimuli before trial onset. At the beginning of a trial, participants were virtually always fixating the position where the picture–word stimulus would come up.

A naming response was considered to be invalid when it included a speech error, when a wrong word was produced, or when the voice key was triggered incorrectly. A manual response was considered to be invalid when the wrong button was pressed. Moreover, a trial was invalid when the vocal response did not have an earlier onset than the manual response (which happened on less than 1% of the trials in all experiments). Error trials were discarded from the analyses of the naming latencies, gaze durations, and manual response latencies.

The vocal response latencies, gaze shift latencies, manual response latencies, and errors were submitted to analyses of variance (ANOVA). The analyses were performed both by participants (F_1) and by items (F_2). In addition, *minF′* (Raaijmakers, Schrijnemakers, & Gremmen, 1999) was computed when both the F_1 and the F_2 reached significance or both were marginally significant (if $.05 < p < .10$). To correct for the difference in absolute latencies between measures (vocal, gaze, and manual), the comparisons of the magnitude of effects between response modalities were performed on standard scores (z) with zero mean and unit standard deviation (Winer, Brown, & Michels, 1991). Standard scores are frequently used to obtain comparability of observations obtained by different measurements. To test for semantic effects, the latencies and errors in the semantic and unrelated conditions were compared. To test for lexicality effects, the unrelated and control conditions were compared. Finally, to test for identity effects, the effect of identical distractors relative to the control condition was assessed. It was determined whether the distractor effects differed between the gaze shifts and the vocal responding and between the gaze shifts and the manual responding. For all comparisons performed by participants and by items, an alpha level of .05 was adopted. Agreement between the distractor effects on the gaze shifting and the vocal responding would indicate that the gaze shifts occurred after vocal response selection. Agreement between the distractor effects on the gaze shifting and the manual responding would suggest that the gaze shifts indexed attention shifts. The comparisons between distractor conditions were also performed for each measure separately (with back-transformation from the standard score units). Tables present distractor condition differences with 95% confidence intervals around the contrasts (Masson & Loftus, 2003) by participants and by items, for the vocal responses, gaze shifts, and manual responses. When the size of the confidence interval is smaller than the size of the contrast, the effect is conventionally significant.

Results and discussion

Fig. 2 displays for each distractor condition the mean latencies for the vocal responses, gaze shifts, and manual responses. The figure shows that for all three measures, the latencies were longer in the semantic than in the unrelated condition, and longer in the unrelated than in the identical condition. For all three measures, the latencies for the identical and control conditions did not differ.

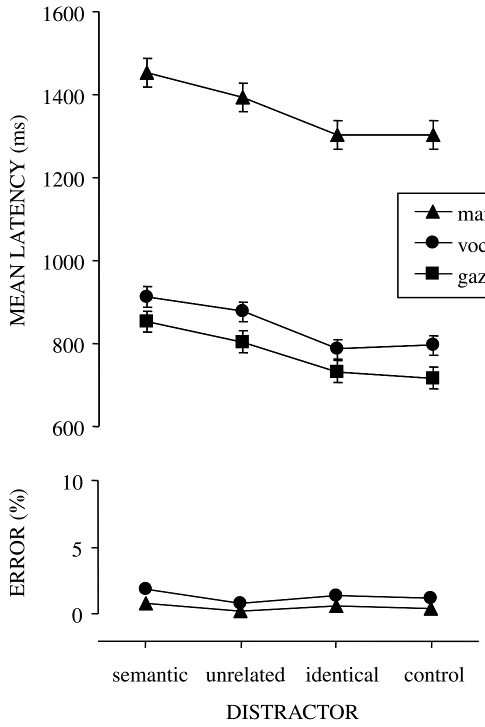

Fig. 2. Mean latencies and error percentages per distractor condition in Experiment 1. The upper panel shows the mean latencies for the vocal picture naming responses, gaze shifts, and manual responses. The error bars indicate the by-participant 95% confidence intervals. The lower panel shows the percentages of erroneous vocal and manual responses.

The statistical analysis of the *z*-scores comparing the latency effects between the vocal responses and gaze shifts yielded an effect of distractor, $F_1(3, 45) = 29.57$, $p = .0001$, $F_2(3, 93) = 27.98$, $p = .0001$, $minF'(3, 124) = 14.38$, $p = .001$, which did not differ between the measures, $F_1(3, 45) = 1.28$, $p = .29$, $F_2(3, 93) = 2.18$, $p = .10$. The analysis of the *z*-scores comparing the latency effects between the gaze shifts and manual responses yielded an effect of distractor, $F_1(3, 45) = 26.19$, $p = .0001$, $F_2(3, 93) = 26.96$, $p = .0001$, $minF'(3, 120) = 13.28$, $p = .001$, which did not reliably differ between the measures, $F_1(3, 45) = 2.29$, $p = .09$, $F_2(3, 93) = 2.33$, $p = .08$. Fig. 2 also shows the error percentages for the vocal and manual responses per distractor condition. The statistical analysis of the errors showed that there was no effect of distractor for the vocal responses, $F_1(3, 45) = 1.14$, $p = .34$, $F_2(3, 93) = 0.62$, $p = .60$, and also not for the manual responses, $F_1(3, 45) = 1.78$, $p = .16$, $F_2(3, 93) = 1.34$, $p = .27$. So, there is no evidence for a speed-accuracy tradeoff in the data.

Table 1 gives the results of the planned comparisons between distractor conditions for each measure separately (vocal, gaze, manual). The latencies were longer in the semantic than in the unrelated condition for all three measures. The latencies were also longer in the unrelated than in the control conditions for all three measures. There was no reliable difference in latency between the identical and control conditions for any of the measures.

Accepting that there is no difference in distractor effects between the measures amounts to accepting the null hypothesis. No difference in distractor effects suggests that there are strong links between the vocal responses and gaze shifts, on the one hand, and between the gaze shifts and manual responses, on the other. The existence of such links implies that there should be relationships between the vocal, gaze, and manual latencies at the stimulus level. This was indeed the case. The Pearson product-moment correlation between the item × condition means of the vocal responses and the gaze shifts was .93, $p = .001$, $N = 128$. This confirms that there was a close relationship between the latency effects for the vocal responses and the gaze shifts. The correlation between the item × condition means of the gaze shifts and the manual responses was .95, $p = .001$, $N = 128$. This shows that there was also a close relationship between the latency effects for the gaze shifts and the manual responses.

To summarize, the magnitudes of the semantic effect (i.e., semantic versus unrelated), the lexicality effect (i.e., unrelated versus control), and the identity effect (i.e., identical versus control) were the same for the vocal

Table 1
Distractor condition differences (in ms) with 95% confidence intervals around the contrasts for the vocal responses, gaze shifts, and manual responses in Experiment 1

Difference	Measure		
	Vocal	Gaze	Manual
Semantic – unrelated	35 (±15, ±25)	49 (±22, ±30)	60 (±24, ±34)
Unrelated – control	81 (±17, ±19)	87 (±16, ±20)	92 (±18, ±21)
Identical – control	−10 (±18, ±13)	15 (±18, ±17)	2 (±26, ±18)

Note. The confidence intervals in parentheses are calculated from the *MSE* for the contrasts in the analysis by participants and by items, respectively.

responses, gaze shifts, and manual responses. The close correspondence between the effects for the vocal responses and gaze shifts is in agreement with the earlier findings suggesting that the signal for moving the eyes occurs after response selection, during the encoding of the form of the vocal word response (Griffin, 2001; Korvorst et al., 2006; Levelt & Meyer, 2000; Meyer et al., 1998, 2003). The data do not provide any evidence that the gaze shifts were triggered before response selection. Moreover, there was a close correspondence between the effects for the gaze shifts and the manual responses, even at the individual item level, suggesting that the gaze shifts indexed attention shifts.

Experiment 2

The second experiment used displays that were identical to those of Experiment 1 except for the control condition. The string of Xs of the control condition of Experiment 1 was replaced by an empty rectangle following Glaser and Düngelhoff (1984). The tasks for the participants were to name the word while ignoring the picture and then to respond to the arrows as in Experiment 1. For example, participants said "vest" in response to the word VEST, while trying to ignore the pictured shirt (semantic), the pictured castle (unrelated), the pictured vest (identical), or an empty rectangle (control). The latencies of the vocal responses, gaze shifts, and manual responses were again recorded.

Method

The method was the same as that of Experiment 1, except that the participants read the words aloud rather than named the pictures of the picture–word stimuli. For the control condition, an empty rectangle was created of 10 × 10 cm.

Results and discussion

Fig. 3 displays for each distractor condition the mean latencies for the vocal responses, gaze shifts, and manual responses. The figure shows that the vocal response latencies did not differ among the distractor conditions. However, for both the gaze shifts and the manual responses, latencies were longer (about 30 ms) with the actual picture distractors (i.e., in the semantic, unrelated, and identical conditions) than with the control rectangles. For both the gaze shifts and manual responses, there were no differences among the conditions with the actual pictures.

The statistical analysis of the *z*-scores comparing the latency effects between the vocal responses and the gaze shifts yielded an effect of distractor in the analyses by participants and items, $F_1(3, 45) = 3.51$, $p = .02$, $F_2(3, 93) = 4.01$, $p = .01$, but not in their combination, $minF'(3, 115) = 1.87$, $p = .14$. Similarly, there was an interaction between distractor and measure in the analyses by participants and items, $F_1(3, 45) = 2.75$, $p = .05$, $F_2(3, 93) = 2.71$, $p = .05$, but not in their combination, $minF'(3, 122) = 1.36$, $p = .26$. Statistical comparisons revealed that the semantic effect (semantic vs. unrelated) did not differ between measures, but the lexicality effect (unrelated vs. control) and the identity effect (identical vs. control) were larger for the gaze shifts than the vocal responses. The analysis of the *z*-scores comparing the latency effects between the gaze shifts and the manual responses yielded an effect of distractor, $F_1(3, 45) = 7.37$, $p = .001$, $F_2(3, 93) = 5.93$, $p = .001$, $minF'(3, 130) = 3.29$, $p = .02$, which did not differ between the measures, $F_1(3,45) = 1.71$, $p = .18$, $F_2(3,93) = 1.26$, $p = .29$. Fig. 3 also shows the error percentages for the vocal and manual responses per distractor condition. The statistical analysis of the errors showed that there was no effect of distractor for the vocal responses, $F_1(3, 45) = 1.60$, $p = .20$, $F_2(3, 93) = 0.89$, $p = .45$, and also

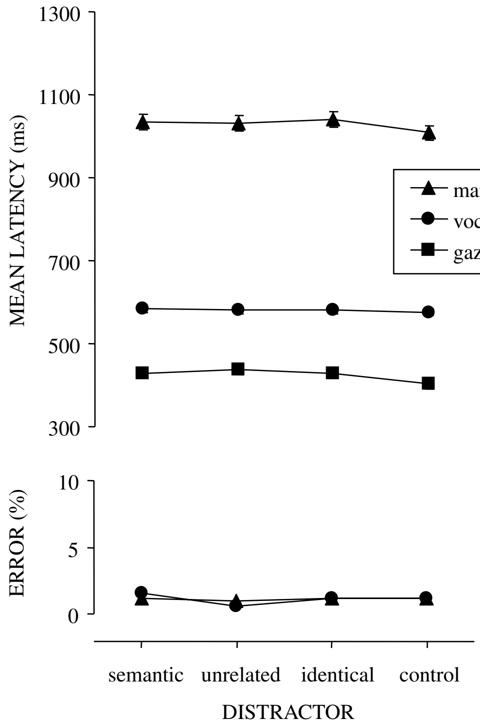

Fig. 3. Mean latencies and error percentages per distractor condition in Experiment 2. The upper panel shows the mean latencies for the vocal word reading responses, gaze shifts, and manual responses. The error bars indicate the by-participant 95% confidence intervals (for the vocal responses and the gaze shifts, the intervals are too small to be visible in the figure). The lower panel shows the percentages of erroneous vocal and manual responses.

Table 2
Distractor condition differences (in ms) with 95% confidence intervals around the contrasts for the vocal responses, gaze shifts, and manual responses in Experiment 2

Difference	Measure		
	Vocal	Gaze	Manual
Semantic – unrelated	2 (±7, ±8)	−8 (±6, ±9)	3 (±11, ±15)
Unrelated – control	5 (±10, ±7)	34 (±11, ±10)	24 (±15, ±12)
Identical – control	4 (±6, ±9)	26 (±11, ±11)	32 (±18, ±16)

Note. The confidence intervals in parentheses are calculated from the *MSE* for the contrasts in the analysis by participants and by items, respectively.

not for the manual responses, $F_1(3, 45) = 0.64$, $p = .59$, $F_2(3, 93) = 0.80$, $p = .50$. So, there is no evidence for a speed-accuracy tradeoff in the data.

Table 2 gives the results of the planned comparisons between distractor conditions for each measure separately (vocal, gaze, manual). The comparisons showed that there was no latency difference between the semantic and unrelated conditions for any of the three measures. The latencies did not differ between the unrelated and the control condition for the vocal responses, but they differed for the gaze shifts and the manual responses. Moreover, the latencies did not differ between the identical and the control condition for the vocal responses, but they differed for the gaze shifts and the manual responses. Thus, the effects for the gaze shifts (i.e., longer latencies for actual pictures than control rectangles) do not fully correspond to the effects for the vocal responses (no differences between conditions). In contrast, the effects for the manual responses agreed with the effects for the gaze shifts. There was a correlation between the item × condition means of the gaze shifts and the manual responses, $r = .51$, $p = .001$, $N = 128$. This confirms that there was a relationship between the latency effects for the gaze shifts and the manual responses. The correlation is smaller than in Experiment 1, presumably because the magnitude of the effects in the present experiment was much smaller than in Experiment 1.

To summarize, there were no distractor effects on the vocal response latencies, but the gaze shift latencies revealed a difference in effect between actual pictures and control stimuli. In word reading, the eyes fixated control stimuli for a shorter time than would be expected on the basis of the vocal response latencies. There was a close correspondence between the effects for the gaze shifts and the manual responses, as observed in Experiment 1. This suggests that the gaze shifts indexed attention shifts.

Comparison of Experiments 1 and 2

The words that were produced as responses in picture naming (Experiment 1) and word reading (Experiment 2) were exactly the same. However, the tasks differed in attention demands, as indexed by the difference in interference effects. The presence of interference effects in picture naming and the absence of such effects on the vocal latencies in word reading suggests that the attention demands were lower for word reading than for picture naming, in agreement with what earlier research suggests (cf. MacLeod, 1991). The participants began to talk after they shifted gaze to the arrows in both experiments. However, the overall difference in onset latency between the vocal responses and the gaze shifts was 66 ms for the picture naming task in Experiment 1, whereas it was 156 ms for the word reading task in Experiment 2. The difference in gaze-to-speech lag between tasks was significant, $F_1(1, 30) = 5.10$, $p = .03$, $F_2(1, 62) = 134.02$, $p = .001$, $minF'(1, 32) = 4.91$, $p = .03$. This demonstrates that the temporal coordination between vocal response planning and gaze shifting may depend on the task. It seems that gaze shifts happen earlier when the attention demands of the task are low than when they are high, as is the case with word reading compared to picture naming.

Experiment 3

In Experiment 2, participants responded to words while trying to ignore picture distractors. The picture distractors did not affect the vocal response latencies, in agreement with earlier research (e.g., Glaser & Düngelhoff, 1984). However, the pictures influenced the gaze durations. Experiment 3 used Experiment 2's displays, and the vocal task was changed from word naming to word categorizing. That is, participants responded to the words by categorizing them (i.e., producing hyperonyms) while trying to ignore the picture distractors. For example, participants said "clothing" in response to the word VEST, while trying to ignore the pictured shirt (semantic), the pictured castle (unrelated), the pictured vest (identical), or the empty rectangle (control). There were now 8 rather than 32 different response words. Earlier research showed that word categorization latencies are affected by picture distractors. In particular, the response latencies with picture distractors are shorter in the semantic than in the unrelated condition, whereas the latencies in the semantic, identical, and control conditions do not differ (e.g., Glaser & Düngelhoff, 1984). The latencies of the vocal responses, gaze shifts, and manual responses were again measured.

Method

The method was the same as that of Experiment 2, except that the participants were producing the hyperonyms of the words rather than reading the words. The Appendix A lists for each word the hyperonym.

Results and discussion

Fig. 4 displays for each distractor condition the mean latencies for the vocal responses, gaze shifts, and manual responses. The figure shows that for all three measures, latencies were shorter in the semantic than in the unrelated condition (indicating semantic facilitation), and longer in the unrelated than in the identical and control conditions. For all three measures, the latencies did not differ between the identical and control conditions. Moreover, the figure shows that the semantic facilitation effect was larger for the vocal responses than for the gaze shifting.

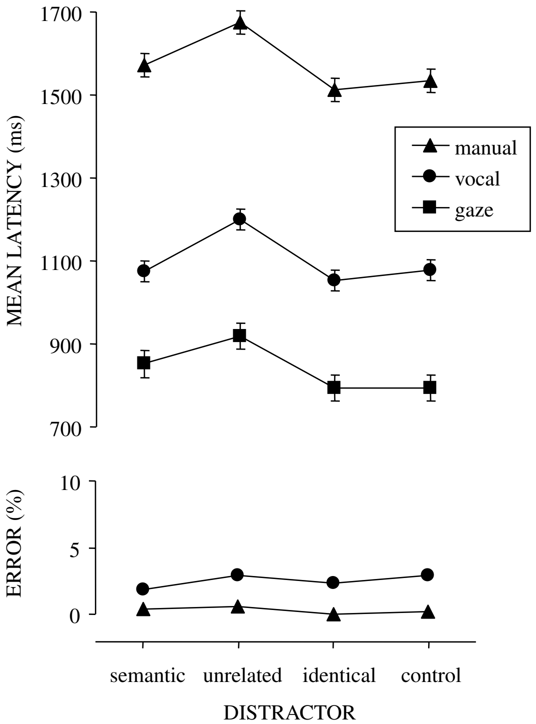

Fig. 4. Mean latencies and error percentages per distractor condition in Experiment 3. The upper panel shows the mean latencies for the vocal word categorizing responses, gaze shifts, and manual responses. The error bars indicate the by-participant 95% confidence intervals. The lower panel shows the percentages of erroneous vocal and manual responses.

The statistical analysis of the *z*-scores comparing the latency effects between the vocal responses and gaze shifts yielded an effect of distractor, $F_1(3, 45) = 23.56$, $p = .0001$, $F_2(3, 93) = 24.83$, $p = .0001$, $minF'(3, 119) = 12.09$, $p = .001$. There was a significant interaction between distractor and measure in the analyses by participants and items, $F_1(3, 45) = 4.50$, $p = .008$, $F_2(3, 93) = 4.29$, $p = .007$, but only a marginally significant effect in their combination, $minF'(3, 123) = 2.20$, $p = .09$. Statistical comparisons revealed that the semantic effect (semantic vs. unrelated) was larger in the vocal responses than in the gaze shifts, whereas the lexicality effect (unrelated vs. control) and identity effect (identical vs. control) did not differ between measures. The analysis of the *z*-scores comparing the latency effects between the gaze shifts and manual responses yielded an effect of distractor, $F_1(3, 45) = 21.32$, $p = .0001$, $F_2(3, 93) = 23.98$, $p = .0001$, $minF'(3, 116) = 11.29$, $p = .001$, which did not reliably differ between the measures, $F_1(3, 45) = 1.55$, $p = .21$, $F_2(3, 93) = 2.37$, $p = .08$. Fig. 4 also shows the error percentages for the vocal and manual responses per distractor condition. The statistical analysis of the errors showed that there was no effect of distractor for the vocal responses, $F_1(3, 45) = 0.90$, $p = .45$, $F_2(3, 93) = 0.76$, $p = .52$, and also not for the manual responses, $F_1(3, 45) = 0.90$, $p = .45$, $F_2(3, 93) = 0.72$, $p = .55$. So, there is no evidence for a speed-accuracy tradeoff in the data.

Table 3 gives the results of the planned comparisons between distractor conditions for each measure separately (vocal, gaze, manual). The comparisons showed that there were latency differences between the semantic and unrelated conditions for all three measures. The latencies were longer in the unrelated than in the control condition for all three measures. There was no reliable difference in latencies between the identical and control conditions for any of the measures.

Thus, there is an agreement between the latency effects for the vocal responses and the gaze shifts except for the semantic condition. The latency difference between the semantic and unrelated conditions is much smaller for the gaze shifts than for the vocal responses. In contrast, the effects for the manual responses correspond with the effects for the gaze shifts. The Pearson correlation between the item × condition means of the gaze shifts and the manual responses was .87, $p = .001$, $N = 128$. This confirms that there was a relationship at the stimulus level between the latency effects for the gaze shifts and the manual responses.

Unlike what was observed in Experiment 2, there was a difference in effect of the unrelated distractor pictures relative to the control rectangles for both the vocal responses and the gaze shifts. Moreover, the semantic

Table 3
Distractor condition differences (in ms) with 95% confidence intervals around the contrasts for the vocal responses, gaze shifts, and manual responses in Experiment 3

Difference	Measure		
	Vocal	Gaze	Manual
Semantic – unrelated	−123 (±19, ±21)	−66 (±32, ±19)	−103 (±21, ±28)
Unrelated – control	121 (±23, ±24)	124 (±38, ±24)	141 (±29, ±30)
Identical – control	−24 (±27, ±27)	0 (±24, ±20)	−20 (±30, ±22)

Note. The confidence intervals in parentheses are calculated from the *MSE* for the contrasts in the analysis by participants and by items, respectively.

effect (i.e., the difference between semantically related and unrelated pictures) was about two times larger for the vocal categorizing responses than for the gaze shifts. Given that the difference between the unrelated and control conditions was the same for the vocal responses and gaze shifts, the data suggest that participants looked much longer at the semantic stimuli than would be expected on the basis of the vocal response latencies. As in Experiments 1 and 2, there was a close correspondence between the effects for the gaze shifts and the manual responses. This suggests, again, that the gaze shifts indexed attention shifts.

Comparison of Experiments 1–3

The stimulus–response compatibility of word categorizing (Experiment 3) and picture naming (Experiment 1) was lower than that of word reading (Experiment 2). Moreover, the number of responses was larger for picture naming and word reading than for word categorizing, namely 32 vs. 8, respectively. Although participants began to talk after they shifted gaze to the arrows in picture naming, word reading, and word categorizing, the overall difference in onset latency between the vocal responses and the gaze shifts was different between tasks. The gaze-to-speech lag was 66 ms for picture naming in Experiment 1, 156 ms for word reading in Experiment 2, and 261 ms for word categorizing in Experiment 3. This suggests that the temporal coordination between vocal response planning and gaze shifting depends on the task. The difference in gaze-to-speech lag between word categorizing and word reading was significant, $F_1(1, 30) = 5.78$, $p = .02$, $F_2(1, 62) = 69.16$, $p = .001$, $minF'(1, 35) = 5.33$, $p = .03$, and the difference in gaze-to-speech lag between word categorizing and picture naming was also significant, $F_1(1, 30) = 15.64$, $p = .001$, $F_2(1, 62) = 241.76$, $p = .001$, $minF'(1,34) = 14.69$, $p = .001$. This demonstrates, again, that the temporal coordination between vocal response planning and gaze shifting may depend on the task. It seems that gaze shifts occur earlier (relative to articulation onset) when the number of responses is small (Experiment 3) than when it is large (Experiments 1 and 2).

General discussion

The research reported in this article addressed the question how eye movements are coordinated between vocal and manual tasks involving visual stimuli in different spatial positions. In particular, when are saccadic eye movements between task stimuli initiated? For task stimuli in different locations requiring vocal and manual responses, it has been assumed that the signal to move the eyes is given before response selection (Meyer & Kieras, 1997a, 1997b). However, evidence suggests that the signal occurs during programming the phonological form of the first response when both stimuli require vocal responses (Meyer et al., 2003). As noted, there are several differences between the task situation modeled by Meyer and Kieras (1997a, 1997b) and the language production research (e.g., Meyer et al., 2003). First, when Tasks 1 and 2 both require a vocal response, the effector system is shared between tasks, whereas with vocal and manual responses it is not. By adopting a serial strategy in planning two vocal responses, the buffering of upcoming Task 2 vocal responses can be limited (Levelt & Meyer, 2000). Second, temporal overlap in planning vocal Tasks 1 and 2 responses may lead to verbal interference. Interference may be avoided by adopting a serial strategy (Meyer & Van der Meulen, 2000; Zelinsky & Murphy, 2000). Third, the task demands differed between experimental situations. There were two vocal responses in the study modeled by Meyer and Kieras (1997a), whereas there were 32 different vocal responses in the experiments of Meyer et al. (2003). If gaze shifts depend on the attention demands of a task, earlier shifts are expected with low than with high demands.

In order to examine when the signal occurs for shifting gaze between stimuli requiring vocal and manual responses, three experiments were run. Speakers were presented with picture–word stimuli. They named the picture (Experiment 1), named the word (Experiment 2), or categorized the word (Experiment 3) and shifted their gaze to a left- or right-pointing arrow to manually indicate its direction. The latencies of the vocal responses, gaze shifts, and manual responses were measured.

As concerns the distractor effects, the results for the vocal response latencies in Experiments 1–3 replicated earlier findings in the literature (e.g., Glaser & Düngelhoff, 1984; Lupker, 1979; Rayner & Springer, 1986; Roelofs, 2003; Smith & Magee, 1980). The latencies of the gaze shifts generally paralleled the vocal response latencies. This suggests that the gaze shifts were initiated after response selection, unlike what Meyer and Kieras (1997a, 1997b) assumed. Importantly, the gaze shifts were initiated after vocal response selection even though Task 2 required a manual response. This suggests that the avoidance of vocal response buffering and the prevention of interference from the second vocal response are not the only reasons for a late gaze shift. Rather, the completion of phonological encoding of the vocal response would seem to be a major determinant of gaze shifts. However, in Experiments 2 and 3, some of the distractor effects were reflected in the gaze shifts but not in the vocal response latencies. Moreover, the difference in onset between vocal responding and gaze shifting depended on the vocal task. The difference in gaze-to-speech lag between vocal tasks and between distractor conditions indicates that the signal to move the eyes is not simply the completion of a predetermined aspect of the programming of the vocal response, unlike what Levelt and Meyer (2000) suggested. In all three experiments, the manual response latencies paralleled the gaze shift latencies, even at an individual item basis. This suggests that the gaze shifts indexed shifts in the focus of attention in performing the tasks.

The trigger for shifting gaze

When exactly did the gaze shifts happen during the course of the three Task 1 processes (picture naming, word reading, word categorizing)? To answer this question, it is important to know what is triggered, the programming and subsequent execution of eye movements or the execution of an already planned saccade. Moreover, estimates of the durations of the component processes of vocal response planning are needed.

As concern the programming and execution of eye movements, evidence suggests that separate processes are involved in computing where and when to move the eyes (e.g., Aslin & Shea, 1987; Becker & Jürgens, 1979). In line with this evidence, prominent models of the control of eye movements in reading assume that word recognition and planning a saccade to the next word happen in parallel (e.g., Reichle, Pollatsek, Fisher, & Rayner, 1998; Reichle, Rayner, & Pollatsek, 2003). For example, in the model proposed by Reichle and colleagues (Reichle et al., 1998, 2003), the oculomotor system is signaled to plan a saccade to the next word when a first stage of word recognition has been completed. Planning the saccade happens in parallel with a second stage of word recognition. The findings of Meyer et al. (2003) suggest that also for picture naming, planning the picture name and planning a saccade to the next object happen in parallel, as I argue next.

It has been estimated that word-form encoding in picture naming takes about 350 ms if the total naming latency is about 600 ms (Indefrey & Levelt, 2004). The process of word-form encoding can be divided into morphological encoding, phonological encoding, and phonetic encoding (Levelt et al., 1999; Roelofs, 1997). Estimates for the durations of these component processes are 80, 125, and 145 ms, respectively (Indefrey & Levelt, 2004). It has been estimated that planning a saccade takes about 150 ms (Meyer & Kieras, 1997a; Reichle et al., 1998). The effect of the length of the picture name (one vs. two syllables) on gaze durations obtained by Meyer et al. (2003) suggests that the gaze shifts occurred during phonological encoding. The average latency in the study of Meyer et al. (2003) was 702 ms for the vocal responses and 448 ms for the gaze shifts (in the pure condition of Experiment 4). Thus, in that study, the gaze shift occurred 254 ms before the onset of articulation. Indefrey and Levelt (2004) based their estimates on a total picture naming latency of 600 ms, whereas the latency was 702 ms in the study by Meyer et al. (2003). Proportionally scaling the estimates of Indefrey and Levelt (2004) for the study of Meyer et al. (2003) yields estimates of 94, 146, and 170 ms for morphological encoding, phonological, and phonetic encoding, respectively. If the completion of phonological encoding triggered the *planning* of a saccade, the difference in onset latency between articulation and gaze shifting should be about 20 ms (i.e., 170 ms for phonetic encoding minus 150 ms for saccade planning). This differs much from the empirically observed difference of 254 ms (Meyer et al., 2003). In contrast, if the completion of phonological encoding triggered the *execution* of an already planned saccade, the onset difference between articulation and gaze shifting should be about 170 ms (i.e., the duration of phonetic encoding), which is much closer to the empirically observed difference (i.e., 254 ms). The empirically observed gaze-to-speech lag suggests that a prepared saccade was triggered even before the completion of phonological encoding, namely *during* the phonological encoding process itself. Fig. 5 illustrates the time line for the onsets of the word-form encoding stages (i.e., morphological, phonological, and phonetic encoding), gaze shifting, and articulation in the study of Meyer et al. (2003). The onsets of the word-form encoding stages are estimates based on Indefrey and Levelt (2004) and the onsets of gaze shifting and articulation were observed by Meyer et al. (2003). To conclude, the data indicate that planning the picture name and planning a saccade happen in parallel, in line with what is assumed for reading (e.g., Reichle et al., 1998, 2003).

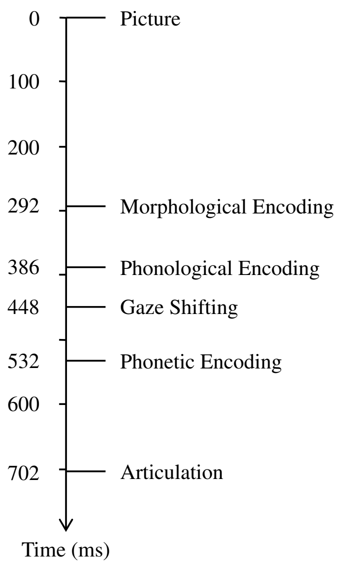

Fig. 5. Time line for the onsets of the word-form encoding stages (i.e., morphological, phonological, and phonetic encoding), gaze shifting, and articulation in the study of Meyer et al. (2003).

In the present Experiment 1, the participants began to articulate the picture name after they shifted gaze to the arrows. The overall difference in onset latency between the vocal responses and the gaze shifts was 66 ms. The mean picture-naming latency in the control condition was 796 ms. Given the estimates of Indefrey and Levelt (2004) for the processes of morphological encoding, phonological encoding, and phonetic encoding, the empirically observed difference of 66 ms in onset latency between vocal picture naming and gaze shifting suggests that the saccade was initiated during the process of phonetic encoding of the picture name. In Experiment 2, the participants began to pronounce the word after they shifted gaze to the arrows. The overall difference in onset latency between the vocal responses and the gaze shifts was 156 ms. The mean word-reading latency in the control condition was 576 ms. Evidence suggests that word-form encoding processes are shared between picture naming and word reading (e.g., Indefrey & Levelt, 2004; Roelofs, 2003, 2004). Given the estimated durations of the component processes of word-form encoding, the empirically observed difference in onset latency between vocal reading and gaze shifting of 156 ms suggests that the saccade was released around the start of phonetic encoding of the word. Finally, in Experiment 3, the participants also began to pronounce the word after they shifted gaze to the arrows. The mean word-categorizing latency in the control condition was 1077 ms. Presumably, the response latency is so long because of the conceptual processes needed to determine the superordinate concept for the word, whereas the word-form encoding latencies are similar to those for picture naming and word reading. If so, a difference in onset between articulation and gaze shifting of 261 ms implies that the signal to move the eyes was given around the start of phonological encoding. This corresponds with the finding that the distractor effects observed for the vocal categorizing latencies (presumed to be lemma retrieval effects, e.g., Roelofs, 1992, 2003) were largely reflected in the gaze shifts.

Modeling the coordination of vocal responding, gaze shifting, and manual responding

The findings in the literature and those from the present Experiments 1–3 may be explained by the following model of the coordination of vocal responding, gaze shifting, and manual responding. The model is illustrated in Fig. 6. The attentional control process works similar to the executive control processes in EPIC-SRD of Meyer and Kieras (1997a, 1997b). It is assumed that in order to maintain acceptable levels of speed and accuracy, participants set a criterion for when the attention shift between the vocal and manual tasks should occur (a similar proposal has been made for the initiation of responding in single-task performance by Lupker, Brown, & Colombo, 1997, and Meyer et al., 2003). In EPIC-SRD, reaching the shift criterion is called the occurrence of the "Task 1 unlocking event" (Meyer & Kieras, 1997b, p. 753), which unlocks Task 2. Presumably, the position of the criterion within the Task 1 process is determined on the basis of the initial trials of an experiment, when participants become familiar with the experimental situation, and the criterion stays more or less constant throughout the experiment (though there may be some dynamic updating during the course of a trial, as discussed later). The placement of the criterion may be determined by a number of factors, with an important one being the need to protect the Task 1 word planning process against distractor interference (cf. Allport, 1980a, 1980b, 1987, 1989; Roelofs, 1997, 2003).

At the beginning of each trial, the attentional control process enables both tasks, engages on Task 1, directs gaze towards the Task 1 stimulus, and maintains engagement on Task 1 and monitors Task 1 performance until the task process reaches the shift criterion. During the planning of the target word, a saccade to the arrow is prepared. The oculomotor system may use low-spatial frequency information to determine the position of the arrow or the position may be retrieved from memory. When the shift criterion is reached during the course of Task 1, attention disengages from Task 1 and shifts to Task 2, directly followed by a signal to the saccadic control system to execute the prepared saccade to the Task 2 stimulus. EPIC-SRD may accommodate this proposal about gaze shifting by assuming that the *planning* of the saccade to the Task 2 stimulus is issued early,

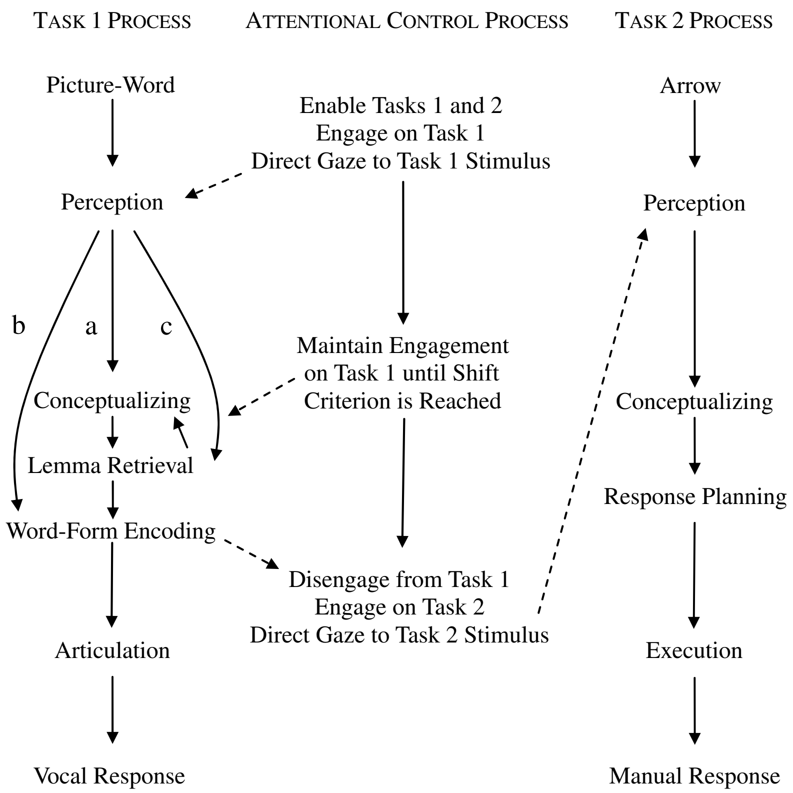

Fig. 6. A model of the coordination of vocal responding, gaze shifting, and manual responding, which combines assumptions of EPIC-SRD (Meyer & Kieras, 1997a, 1997b) and WEAVER++ (Roelofs, 1992, 2003). According to WEAVER++, picture naming involves picture perception, conceptualizing, lemma retrieval, word-form encoding, and articulation (route a); word reading involves word perception, word-form encoding, and articulation (route b); and word categorizing involves word perception, lemma retrieval (for the word), conceptualizing (determining the superordinate concept), lemma retrieval (for the superordinate term), word-form encoding, and articulation (route c).

but that the signal to *execute* the prepared saccade (i.e., to move the eyes) depends on having reached the Task 1 unlocking event.

According to the WEAVER++ model (Roelofs, 1992, 2003), picture naming involves picture perception, conceptualizing, lemma retrieval, word-form encoding, and articulation (route a in Fig. 6). Word reading involves word perception, word-form encoding, and articulation (route b). Word categorizing involves word perception, lemma retrieval (for the word), conceptualizing (determining the superordinate concept), lemma retrieval (for the superordinate term), word-form encoding, and articulation (route c). Words are planned by spreading activation through a lexical network and application of condition-action production rules. In securing task-relevant control, the model uses goal symbols, whose presence constitutes one of the conditions for the firing of a production rule. When a goal symbol is placed in working memory, the attention of the system is focussed on those production rules that include this goal among their conditions. The EPIC-SRD model of dual-task performance advanced by Meyer and Kieras (1997a, 1997b) also uses goal-factored production rules to achieve attentional control. EPIC's executive production rules coordinate cognitive processes across the tasks in a dual-task situation. The processes illustrated in Fig. 6 may be computationally implemented by combining WEAVER++'s word planning and attentional control processes and EPIC's task coordination processes. In particular, production rules are needed for goal setting and shifting and for directing gaze (cf. Roelofs, 2003). Later in this article, the utility of this approach is illustrated by a simulation of Experiment 3 using the model shown in Fig. 6.

The model outlined in Fig. 6 explains the difference in results between earlier studies (e.g., between Meyer & Kieras, 1997a, 1997b, and Meyer et al., 1998) in terms of differences in shift criteria between tasks. Presumably, the shift criterion is set earlier for tasks with low complexity of response selection and low attention demands (e.g., Meyer & Kieras, 1997a, 1997b) than for tasks with high selection complexity and high attention demands (e.g., Meyer et al., 2003). The reported Experiments 1–3 differed in the timing of gaze shifts relative to the onset of articulation. Relative to articulation onset, gaze shifts occurred later for picture naming (Experiment 1) than for word reading (Experiment 2), and later for word reading than for word categorizing (Experiment 3). In terms of the model, the shift criterion was set earlier for word categorizing than for word reading, and earlier for word reading than for picture naming. The differences in shift criteria reflect the difference in need to protect the word planning process in picture naming, word reading, and word categorizing against distractor words and pictures.

In picture naming in Experiment 1, attentional control of Task 1 was needed for a relatively long time in order to protect the planning of the picture name against the distractor word (cf. Roelofs, 2003). Because the distractor word activates its output word form, planning the picture name requires attention until word-form encoding has almost been completed. This may explain why the shift criterion was set so late during the Task 1 process (i.e., during phonetic encoding). Moreover, it explains why the difference in onset between vocal responding and gaze shifting was smaller in Experiment 1 than in the experiment of Meyer et al. (2003), 66 vs. 254 ms. In the latter experiment, there were no distractor words, so the word-form encoding in picture naming did not have to be protected against distractor words, and attention and gaze could shift earlier in time. Attentional control of Task 1 was also needed for a much shorter time for word reading in Experiment 2 than for picture naming in Experiment 1. Basically, attention was needed for input selection (i.e., selecting the word for further processing). Distractor pictures do not activate (much) their word forms (Roelofs, 2003, 2006), thus attention is not required to protect the word-form encoding process against the distractor, unlike what was the case with picture naming in Experiment 1. This may explain why the shift criterion was set earlier in Experiment 2 than in Experiment 1. Similarly, attention to word categorizing in Experiment 3 was needed for input selection (selecting the word for further processing) and for the conceptualization processes (determining the superordinate concept). These processes need to be protected against the distractor pictures. However, once the response is selected (the superordinate term for the word), encoding the word form does not need to be protected against the distractor pictures, because the pictures do not activate (much) their word forms. Thus, again, attention and gaze may shift well before the completion of word-form encoding, as in Experiment 2.

The differences in shift criterion explain the differences in the onset of articulation and gaze shifting between tasks (picture naming, word reading, and word categorizing). However, task-dependent differences in shift criterion do not explain why the timing of gaze shifts may depend, to some extent, on the distractor condition. The latter may happen if the shift criterion is updated during the course of a trial, dependent on the experience with the distractor. In EPIC-SRD a distinction is made between "static" and "progressive" unlocking of Task 2. With static unlocking, the Task 1 unlocking event remains fixed throughout each trial, whereas with progressive unlocking, "the specification of the Task 1 unlocking event is contingently updated during the course of each trial" (Meyer & Kieras, 1997b, p. 761).

In Experiment 2, participants had to read aloud the word and to ignore the picture of the picture–word stimuli. Participants moved their eyes from words with rectangles around them a little sooner than they moved their eyes from words with actual pictures around them, even though the vocal response latencies did not differ among conditions. It may be that the attentional control process dynamically adapted the shift criterion, such that it was set later for the actual pictures than for the rectangles, because the potential for interference was greater with the actual pictures than the rectangles. Consequently, speakers moved their eyes away sooner with the rectangles than the actual pictures. In Experiment 3, participants had to categorize the word and to ignore the picture of the picture–word stimuli. Vocal word categorizing was slowed by unrelated picture distractors relative to the other distractors, and the gaze shifts followed this pattern. Relative to the control condition, the effect of unrelated and identical picture distractors was now the same for the vocal responses and the gaze shifts. This is different from what was observed in Experiment 2. Moreover, with word categorizing, the difference in effect between semantically related and unrelated pictures was much smaller for the gaze shifts than for the vocal responses. Participants fixated the picture–word stimuli in the semantic condition much longer than would be expected on the basis of the vocal categorizing latencies. It may be that the attentional control process postponed the Task 2 unlocking in the semantic condition relative to the other conditions, perhaps because the semantic distractors confusingly provided both congruent and incongruent information. Consequently, speakers moved their eyes away later in the semantic condition than in the other conditions.

Computer simulations of Experiment 3 using the model outlined in Fig. 6 demonstrated the utility of the theoretical approach. The model combines WEAVER++ (Roelofs, 1992, 2003) and assumptions of the strategic response-deferment (SRD) model of Meyer and Kieras (1997a, 1997b). The vocal latencies were obtained by running word categorizing with picture distractors in WEAVER++ and adding a constant residual latency of 775 ms for the perceptual and form encoding processes not covered by the model (cf. Luce, 1986). The

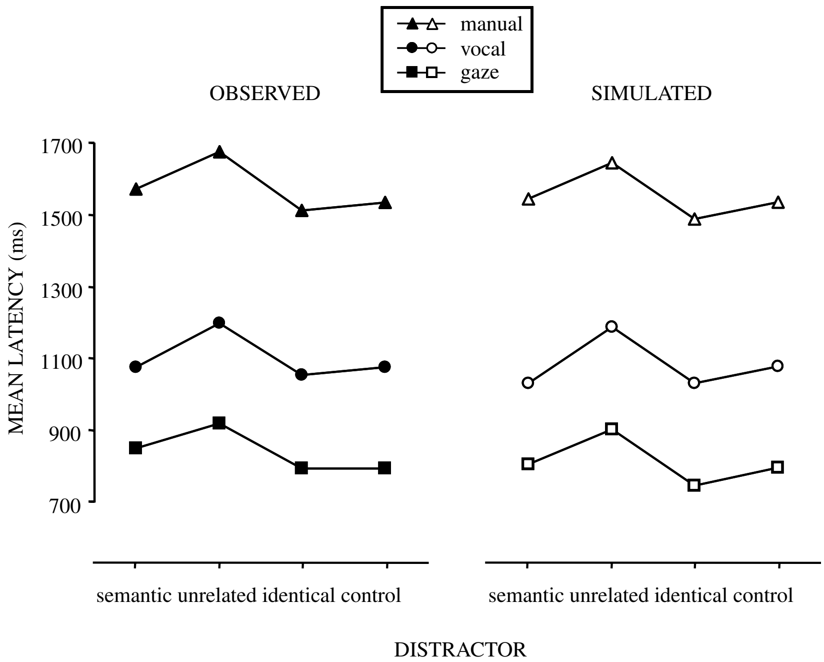

Fig. 7. Simulation of the coordination of vocal responding, gaze shifting, and manual responding in Experiment 3 (word categorizing with picture distractors) using WEAVER++ (Roelofs, 1992, 2003) and assumptions of EPIC-SRD of Meyer and Kieras (1997a). The left-hand panel shows the real data of Experiment 3 and the right-hand panel shows the results of the computer simulations.

WEAVER++ parameters were identical to those of Roelofs (1992). The simulations reported in Roelofs (1992) concerned semantic and unrelated distractors only, whereas the present simulations also included identical and control conditions. The mean gaze-to-speech lag was assumed to be 226 ms for the semantic condition and 283 ms otherwise, reflecting the dynamic updating of the shift criterion during a trial. The mean gaze-to-manual lag was assumed to be 739 ms. This lag covered the duration of the saccade from the picture–word stimulus to the arrow and the duration of the Task 2 process. Estimates of the intervals were obtained from the real data of Experiment 3. Note that the number of parameters (i.e., 4) is much less than the number of data points (i.e., 12), making the fit of the model to the data nontrivial. Fig. 7 shows the simulation results together with the data obtained in Experiment 3. The fit between model and data was good ($R^2 = .99$, meaning that the model accounts for 99% of the systematic variance in the empirical mean latencies). The good model fit demonstrates the utility of the theoretical approach outlined in Fig. 6.

Conclusions

Previous studies found no agreement on when participants shift their eyes from one task stimulus to the next in a multiple-task situation. The present findings suggest that the time it takes to select and program a vocal response is a major determinant of the timing of saccadic eye movements between visual stimuli for primary vocal and secondary manual tasks. However, the exact temporal coordination of vocal responding and gaze shifting depends on the vocal task (picture naming, word reading, and word categorizing). Moreover, the coordination within a task depends, to some extent, on the type of stimulus (i.e., the relationship between picture and word in the present experiments). In all the experiments, there was a close correspondence between the distractor interference and facilitation effects on the gaze shifting and the manual responding, which suggests that the gaze shifts reflected shifts of attentional engagement between the vocal and manual tasks. The results were interpreted as showing that participants strategically set and adjust a criterion for shifting gaze to the secondary task, dependent on variables such as the primary task and the relationship between picture and word. Computer simulations showed that a simple extension of WEAVER++ (Roelofs, 1992, 2003) with assumptions about criterion setting and attentional control in the coordination of vocal responding, gaze shifting, and manual responding quantitatively accounts for the key findings.

Appendix A

Materials of the experiments

Target		Distractor		
Hyperonym	Basic level	Semantic	Unrelated	Identical
dier (animal)	zwaan (swan)	schildpad	rok	zwaan
	schildpad (tortoise)	zwaan	beker	schildpad
	konijn (rabbit)	hert	paleis	konijn
	hert (deer)	konijn	bureau	hert
kleding (clothing)	trui (sweater)	rok	dolk	trui
	rok (skirt)	trui	zwaan	rok
	hemd (shirt)	vest	oor	hemd
	vest (vest)	hemd	kasteel	vest
vervoer (transportation)	fiets (bike)	trein	kast	fiets
	trein (train)	fiets	arm	trein
	schip (ship)	vliegtuig	been	schip
	vliegtuig (plane)	schip	glas	vliegtuig
gebouw (building)	molen (windmill)	fabriek	kom	molen
	fabriek (factory)	molen	neus	fabriek
	kasteel (castle)	paleis	vest	kasteel
	paleis (palace)	kasteel	konijn	paleis
wapen (weapon)	dolk (dagger)	speer	trui	dolk
	speer (spear)	dolk	tafel	speer
	kanon (cannon)	pistool	bord	kanon
	pistool (pistol)	kanon	bed	pistool
servies (service)	beker (cup)	kom	schildpad	beker
	kom (bowl)	beker	molen	kom
	glas (glass)	bord	vliegtuig	glas
	bord (plate)	glas	kanon	bord
meubel (furniture)	tafel (table)	kast	speer	tafel
	kast (cupboard)	tafel	fiets	kast
	bed (bed)	bureau	pistool	bed
	bureau (desk)	bed	hert	bureau
lichaamsdeel (body part)	arm (arm)	neus	trein	arm
	neus (nose)	arm	fabriek	neus
	been (leg)	oor	schip	been
	oor (ear)	been	hemd	oor

References

Allport, D. A. (1980a). Patterns and actions: cognitive mechanisms are content-specific. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 26–64). London: Routledge & Kegan Paul.

Allport, D. A. (1980b). Attention and performance. In G. Claxton (Ed.), *Cognitive psychology: New directions* (pp. 112–153). London: Routledge & Kegan Paul.

Allport, D. A. (1987). Selection for action: some behavioral and neurophysiological considerations of attention and action. In H. Heuer & A. F. Sanders (Eds.), *Perspectives on perception and action* (pp. 395–419). Hillsdale, NJ: LEA.

Allport, D. A. (1989). Visual attention. In M. I. Posner (Ed.), *Foundations of cognitive science* (pp. 631–682). Cambridge, MA: MIT Press.

Aslin, R. N., & Shea, S. L. (1987). The amplitude and angle of saccades to double-step target displacements. *Vision Research, 27*, 1925–1942.

Becker, W., & Jürgens, R. (1979). Analysis of the saccade system by means of double step stimulation. *Vision Research, 19*, 967–983.

Bock, K., Irwin, D. E., Davidson, D. J., & Levelt, W. J. M. (2003). Minding the clock. *Journal of Memory and Language, 48*, 653–685.

Charniak, E., & McDermott, D. (1985). *Introduction to artificial intelligence*. Reading, MA: Addison-Wesley.

Donders, F. C. (1870). Die Bewegungen des Auges, veranschaulicht durch das Phaenophthalmotrop [The movements of the eye, demonstrated through the ophthalmotrope]. *Archiv für Ophthalmologie, 16*, 154–175.

Glaser, W. R., & Düngelhoff, F.-J. (1984). The time course of picture–word interference. *Journal of Experimental Psychology: Human Perception and Performance, 10*, 640–654.

Glaser, W. R., & Glaser, M. O. (1989). Context effects in Stroop-like word and picture processing. *Journal of Experimental Psychology: General, 118*, 13–42.

Griffin, Z. M. (2001). Gaze durations during speech reflect word selection and phonological encoding. *Cognition, 82*, B1–B14.

Griffin, Z. M., & Bock, K. (2000). What the eyes say about speaking. *Psychological Science, 11*, 274–279.

Harley, T. A. (1984). A critique of top-down independent levels models of speech production: Evidence from non-plan-internal speech errors. *Cognitive Science, 8*, 191–219.

Hoffman, J., & Subramaniam, B. (1995). The role of visual attention in saccadic eye movements. *Perception & Psychophysics, 57*, 787–795.

Indefrey, P., & Levelt, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition, 92*, 101–144.

Kornblum, S., Hasbroucq, T., & Osman, A. (1990). Dimensional overlap: cognitive basis for stimulus–response compatibility: a model and taxonomy. *Psychological Review, 97*, 253–270.

Korvorst, M., Roelofs, A., & Levelt, W. J. M. (2006). Incrementality in naming and reading complex numerals: Evidence from eyetracking. *Quarterly Journal of Experimental Psychology, 59*, 296–311.

Kowler, E., Anderson, E., Dosher, B., & Blaser, E. (1995). The role of attention in the programming of saccades. *Vision Research, 35*, 1837–1916.

Levelt, W. J. M., & Meyer, A. S. (2000). Word for word: multiple lexical access in speech production. *European Journal of Cognitive Psychology, 12*, 433–452.

Levelt, W. J. M., Roelofs, A., & Meyer, A. S. (1999). A theory of lexical access in speech production. *Behavioral and Brain Sciences, 22*, 1–38.

Lupker, S. J. (1979). The semantic nature of response competition in the picture–word interference task. *Memory & Cognition, 7*, 485–495.

Lupker, S. J., Brown, P., & Colombo, L. (1997). Strategic control in a naming task: Changing routes or changing deadlines? *Journal of Experimental Psychology: Learning, Memory, and Cognition, 23*, 570–590.

Luce, R. D. (1986). *Response times: Their role in inferring elementary mental organization*. New York: Oxford University Press.

MacLeod, C. M. (1991). Half a century of research on the Stroop effect: an integrative review. *Psychological Bulletin, 109*, 163–203.

Masson, M. E. J., & Loftus, G. R. (2003). Using confidence intervals for graphically based data interpretation. *Canadian Journal of Experimental Psychology, 57*, 203–220.

Meyer, A. S., Roelofs, A., & Levelt, W. J. M. (2003). Word length effects in object naming: the role of a response criterion. *Journal of Memory and Language, 48*, 131–147.

Meyer, A. S., Sleiderink, A. M., & Levelt, W. J. M. (1998). Viewing and naming objects. *Cognition, 66*, B25–B33.

Meyer, A. S., & Van der Meulen, F. F. (2000). Phonological priming of picture viewing and picture naming. *Psychonomic Bulletin & Review, 7*, 314–319.

Meyer, D. E., & Kieras, D. E. (1997a). A computational theory of executive cognitive processes and multiple-task performance: Part 1. Basic mechanisms. *Psychological Review, 104*, 3–65.

Meyer, D. E., & Kieras, D. E. (1997b). A computational theory of executive cognitive processes and multiple-task performance: Part 2. Accounts of psychological refractory-period phenomena. *Psychological Review, 104*, 749–791.

Monsell, S. (1996). Control of mental processes. In V. Bruce (Ed.), *Unresolved mysteries of the mind: Tutorial essays in cognition* (pp. 93–148). Hove, UK: Erlbaum.

Morsella, E., & Miozzo, M. (2002). Evidence for a cascade model of lexical access in speech production. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 28*, 555–563.

Navarrete, E., & Costa, A. (2005). Phonological activation of ignored pictures: further evidence for a cascade model of lexical access. *Journal of Memory and Language, 53*, 359–377.

Pashler, H. (1994). Dual-task interference in simple tasks: data and theory. *Psychological Bulletin, 116*, 220–244.

Posner, M. I. (1978). *Chronometric explorations of mind*. Hillsdale, NJ: Erlbaum.

Posner, M. I. (1980). Orienting of attention. *Quarterly Journal of Experimental Psychology, 32*, 3–25.

Posner, M. I., Snyder, C. R. R., & Davidson, B. J. (1980). Attention and the detection of signals. *Journal of Experimental Psychology: General, 109*, 160–174.

Raaijmakers, J. G. W., Schrijnemakers, J. M. C., & Gremmen, F. (1999). How to deal with "The language-as-fixed-effect fallacy": Common misconceptions and alternative solutions. *Journal of Memory and Language, 41*, 416–426.

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin, 124*, 372–422.

Rayner, K., & Springer, C. J. (1986). Graphemic and semantic similarity effects in the picture–word interference task. *British Journal of Psychology, 77*, 207–222.

Reichle, E. D., Pollatsek, A., Fisher, D. L., & Rayner, K. (1998). Toward a model of eye movement control in reading. *Psychological Review, 105*, 125–157.

Reichle, E. D., Rayner, K., & Pollatsek, A. (2003). The E-Z Reader model of eye-movement control in reading: comparisons to other models. *Behavioral and Brain Sciences, 26*, 445–526.

Roelofs, A. (1992). A spreading-activation theory of lemma retrieval in speaking. *Cognition, 42*, 107–142.

Roelofs, A. (1997). The WEAVER model of word-form encoding in speech production. *Cognition, 64*, 249–284.

Roelofs, A. (2003). Goal-referenced selection of verbal action: modeling attentional control in the Stroop task. *Psychological Review, 110*, 88–125.

Roelofs, A. (2004). Seriality of phonological encoding in naming objects and reading their names. *Memory & Cognition, 32*, 212–222.

Roelofs, A. (2006). Context effects of pictures and words in naming objects, reading words, and generating simple phrases. *Quarterly Journal of Experimental Psychology, 59*, 1764–1784.

Sanders, A. F. (1998). *Elements of human performance: Reaction processes and attention in human skill*. London: LEA.

Sanders, A. F., & Lamers, J. M. (2002). The Eriksen flanker effect revisited. *Acta Psychologica, 109*, 41–56.

Shallice, T. (1988). *From neuropsychology to mental structure*. Cambridge, UK: Cambridge University Press.

Shepherd, M., Findlay, J., & Hockey, R. (1986). The relationship between eye movements and spatial attention. *Quarterly Journal of Experimental Psychology, 38A*, 475–491.

Smith, M. C., & Magee, L. E. (1980). Tracing the time course of picture–word processing. *Journal of Experimental Psychology: General, 109*, 373–392.

Sternberg, S. (1969). The discovery of processing stages: extension of Donders' method. *Acta Psychologica, 30*, 276–315.

Telford, C. W. (1931). The refractory phase of voluntary and associative responses. *Journal of Experimental Psychology, 14*, 1–36.

Tsal, Y. (1983). Movements of attention across the visual field. *Journal of Experimental Psychology: Human Perception and Performance, 9*, 523–530.

Van Duren, L. L., & Sanders, A. F. (1995). Signal processing during and across saccades. *Acta Psychologica, 89*, 121–147.

Welford, A. T. (1952). The "psychological refractory period" and the timing of high-speed performance—A review and a theory. *British Journal of Psychology, 43*, 2–19.

Winer, B. J., Brown, D. R., & Michels, K. M. (1991). *Statistical principles in experimental design*. New York: McGraw-Hill.

Wundt, W. (1862). Über die Bewegung der Augen [On the movement of the eyes]. *Archiv für Ophthalmologie, 8*, 1–87.

Wundt, W. (1897). *Outlines of psychology*. Leipzig, Germany: Engelmann.

Zelinsky, G. J., & Murphy, G. L. (2000). Synchronizing visual and language processing: an effect of object name length on eye movements. *Psychological Science, 11*, 125–131.