Running head: External vs. internal speech monitoring

Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception

Falk Huettig[a] and Robert J. Hartsuiker[b]

[a]Max Planck Institute for Psycholinguistics

[b]Ghent University

Correspondence should be addressed to:

Falk Huettig
Max Planck Institute for Psycholinguistics
P.O. Box 310
6500 AH Nijmegen
The Netherlands
E-mail: falk.huettig@mpi.nl
phone: +31-24-3521374
fax: +31-24-3521213

Abstract

Theories of verbal self-monitoring generally assume an internal (pre-articulatory) monitoring channel, but there is debate about whether this channel relies on speech perception or on production-internal mechanisms. Perception-based theories predict that listening to one's own inner speech has similar behavioral consequences as listening to someone else's speech. Our experiment therefore registered eye-movements while speakers named objects accompanied by phonologically related or unrelated written words. The data showed that listening to one's own speech drives eye-movements to phonologically related words, just as listening to someone else's speech does in perception experiments. The time-course of these eye-movements was very similar to that in other-perception (starting 300 ms post-articulation), which demonstrates that these eye-movements were driven by the perception of overt speech, not inner speech. We conclude that external, but not internal monitoring, is based on speech perception.

Key words: Language production, verbal self-monitoring, perceptual loop theory, speech perception

When we speak we can listen to ourselves, and so we can use speech perception to monitor our own overt speech for deviations from plan. Theories of speech monitoring (e.g., Hartsuiker & Kolk, 2001; Levelt, 1989; Postma, 2000) generally assume that speakers can exploit multiple monitoring channels: In addition to monitoring their overt speech by listening to it (the external channel), they can also inspect an internal representation of speech before articulation (the internal channel). It is unclear, however, how they monitor this internal representation. This article asks whether internal-channel monitoring engages speech perception (e.g., Levelt, 1989) or engages language production internal devices (e.g., Laver, 1980; see Postma, 2000 for review).

The existence of an internal monitoring channel is supported by many findings. For example, speakers can detect their speech errors in the absence of external-channel monitoring, namely when overt speech is noise-masked (Lackner & Tuller, 1979). Speakers can also detect errors in their silent recitation of tongue twisters, with a very similar distribution of error types to that in overt recitation (Dell & Repka, 1992; Oppenheim & Dell, 2008). Additionally, erroneous words are sometimes interrupted very early on (after only one phoneme, as in *v-horizontal*, Levelt, 1983). This makes it unlikely that the error had been detected in overt speech, because overt speech-error detection involves the time-consuming processes of speech perception, checking, and self-interruption. Finally, speakers are less likely to produce a slip of the tongue when this results in a taboo utterance (*tool kits => cool tits*) than a neutral utterance (*tool carts => cool tarts*), but they show an increased galvanic skin response when correctly saying the taboo-inducing words (*tool* *kits*) (Motley, Camden, & Baars, 1982). The latter finding suggests that the taboo-slip was made internally (leading to an emotional response), but detected via the internal channel and corrected.

What, however, is the nature of internal-channel monitoring? According to Levelt's (1983; 1989) Perceptual Loop Theory the internal channel uses the speech perception system, just as the external channel does. Specifically, the output of phonological encoding would feed directly into the speech perception system; processing then follows the same processing route as overtly perceived speech, and the end result ("parsed speech") feeds into a control component which is responsible for the checking, interruption, and self-correction functions. The computational advantage of this proposal is that the internal monitoring system can exploit cognitive machinery already in place for perceiving overt speech (be it self- or other-produced).

In contrast, other theories postulate a monitoring device or devices within the language production system itself (e.g., Laver, 1980; Nickels & Howard, 1995; Oomen, Postma, & Kolk, 2005). According to such theories, the production system generates an output at each processing level, and detects whether the output is consistent with a target. Mattson and Baars (1992), for example, suggested that a monitor might exploit the activation dynamics at the output layer of units in a connectionist network. When all is going to plan, only the correct representation should be highly activated. But if we (are about to) err, both the correct and erroneous representations will be active, because the target is primed by correct units at the previous layer (see Botvinick, Braver, Barch, Carter, & Cohen, 2001, for a similar proposal in the domain of action monitoring).

Suggestive evidence that self-monitoring is possible without engaging speech comprehension comes from dissociations between monitoring and comprehension in the neuropsychological literature. There are cases of patients with jargon aphasia (Marshall, Robson, Pring, & Chiat, 1998), Parkinson's disease (McNamara, Obler, Au, Durso, & Albert, 1992), and dementia of the Alzheimer type (McNamara et al., 1992) showing impairments in monitoring but with relatively intact comprehension skills. The reverse dissociation has also been reported: Marshall, Rappaport, and Garcia-Bunuel (1985) presented the case of a patient with good monitoring, despite poor comprehension.

Additionally, Nickels and Howard (1995) found no relation between indices of monitoring (number of uncorrected phonological errors, self-interruptions, and self-repairs) and speech comprehension skills in aphasics. Moreover, Oomen et al. (2005) reported one aphasic patient (G.) with comparable deficits in production and monitoring. These studies suggest that, at least in aphasia, internal monitoring uses the production, rather than the comprehension system.

In contrast, Özdemir, Roelofs, and Levelt (2007) presented experimental evidence that appears to support a perceptually-based monitoring system. They argued that if internal monitoring engages perception, monitoring should be sensitive to factors known to influence speech perception. One such factor is the *uniqueness point*: the phoneme at which the word diverges from all other words in the language (e.g., Marslen-Wilson, 1990). In studies with the phoneme monitoring task ("*e.g., push the button if the word you will hear contains a /b/"*) participants respond more quickly when the target phoneme follows the uniqueness point than when it precedes it; for target phonemes following the uniqueness point, latencies are shorter when distance to the uniqueness point is longer. To test whether internal monitoring is also sensitive to the uniqueness point, Özdemir et al. used a phoneme monitoring task in production. Participants viewed line drawings of simple objects (e.g., a *puzzle*), and were asked to silently monitor for a target phoneme (e.g., /l/) in the name of that picture (see also Wheeldon & Levelt, 1995). In a control condition, participants overtly named the picture. There were effects of distance to the uniqueness point in phoneme monitoring (similar to those in the perceptual variant of this task), but not in picture naming, which Özdemir et al. interpreted as evidence for perceptual effects on internal monitoring.

However, this conclusion rests on the assumption that a metalinguistic task like phoneme monitoring engages the same processes that speakers habitually use when monitoring their own speech. There is no evidence to support this assumption, and in fact it has been suggested that speakers can only "listen" to internal speech when performing a silent task (like Özdemir et al.'s phoneme monitoring task), but not when speaking out loud (Vigliocco & Hartsuiker, 2002). We therefore tested for perceptual effects on internal monitoring, but without employing a metalinguistic task.

The current study exploited the phenomenon that speech perception steers overt visual attention, as measured by eye-movements in a visual scene (e.g., Cooper, 1974; Tanenhaus, Spivey-Knowlton, Eberhard & Sedivy, 1995). Crucially, in visual displays that contain written words, listeners tend to fixate more often on words with an overlapping than non-overlapping phonology with the target word (Huettig & McQueen, 2007). For example, given an auditorily presented sentence containing the Dutch word *beker* (beaker), speakers fixated the phonologically related word *bever* (beaver) more often than an unrelated word, or words related on other dimensions (semantic category or shape of referent). Proportions of looks to the phonologically related words started to diverge from those to the other words from around 300 ms after acoustic onset, consistent with estimates that it takes about 200 ms to program and execute a linguistically mediated eye-movement (see Altmann & Kamide, 2004 for further discussion) and allowing some time for the processing of word-initial acoustic information.

To test whether internal monitoring engages speech perception, we conducted an analogous experiment in word production, hypothesizing that hearing ourselves should influence eye-movements similarly as hearing others. Participants viewed displays containing a target object to be named and three written-word distractors. In the critical condition, one of these words was phonologically related to the name of the object and two words were unrelated (Figure 1a). Of interest was the pattern of fixations to the phonologically related distractors prior to and after the onset of overt articulation. A perceptually-based theory of internal monitoring predicts that speech perception is already engaged in the time interval before overt articulation (e.g., Hartsuiker & Kolk, 2001; Indefrey & Levelt, 2004). Estimates of the head-start of inner-speech perception on overt word onset vary between 145 ms (Indefrey & Levelt, 2004) and 250 ms (Hartsuiker & Kolk, 2001; Levelt, 1989). Therefore, the theory that internal monitoring is based on speech perception predicts that an increase in looks to the distractor words begins about 50 ms before word onset given the more liberal estimate of a 250 ms head start and allowing 200 ms to program and execute an eye-movement (-250 ms + 200 ms = - 50 ms). Given the more conservative estimate of a 145 ms head start, this theory predicts an increase in looks to the distractor words starting from 55 ms after word onset (-145 ms + 200 ms = 55 ms). In contrast, if only external monitoring engages speech perception, fixations to the phonological distractor should only be conditional upon the onset of overt speech and should thus begin not before 200 ms (0 ms + 200 ms = 200 ms) post word onset, similar to the time course of such fixations in the perception of other-produced speech.

Insert Figure 1 about here

Similar to a recent production study that measured eye-movements (Huettig & Hartsuiker, 2008), our experiment built in several controls. First, we included a condition with semantically related words (i.e., category coordinates of the target object) to further assess whether listening to oneself and listening to someone else is similar. Huettig and McQueen (2007) found a phonological effect, but no semantic effect with written word displays, in listening to other-produced speech, and we expected to replicate this pattern in listening to self-produced speech. Second, to assess whether the presence of phonologically related and semantically related words under our task circumstances affected the word production process itself, we included a control condition in which the phonologically (semantically) related competitor was replaced by an unrelated item (Figure 1b). We could thus assess picture naming latencies for the same stimuli in the presence or absence of a related word. If the presence of related words affects word production (despite having multiple distractor words and having spatial displacement between target object and distractor words), one would expect the standard pattern of context effects in picture-word interference, namely semantic interference and phonological facilitation (e.g., Schriefers et al., 1990). Third, to assess whether the critical competitor words intrinsically draw visual attention (because of some unknown uncontrolled variable), or whether the mere presence of two related items in one display intrisically draws attention, we included a further control condition (the "named distractor condition") in which the target object was replaced by its written name (and became a distractor) and in which an unrelated distractor was replaced by its image (and hence became a target object; Figure 1c). If any increase in looks to the related competitors in the experimental conditions is a result of some confound, a similar increase in looks to those competitors should be observed in this control condition. For completeness we also included a control condition for the named distractor condition, in which the competitor word was replaced by an unrelated word (Figure 1d).

Method

Participants

Fourty-eight Ghent University students, all native speakers of Dutch, participated in exchange for course credits. All had normal or corrected-to-normal vision.

Stimuli

There were 48 sets of visual displays. Each display contained a target object (the object to be named), and three written distractor words, with one item (object or word) in each corner. The approximate size of each visual object was 8 x 8 cm. The objects/ words were randomly assigned to quadrants of the display. The individual black and white line drawings were taken from the Severens, Van Lommel, Ratinckx, and Hartsuiker (2005) set. In each display, all object names and words started with a different phoneme, except for the phonologically related items. The stimuli in each display were matched on log word frequency of their names. There were four versions of each visual display (Figure 1): an experimental condition (e.g., target object *hart* (*heart),* related competitor word *(harp)*, distractor words, *zetel* (couch) and *raam* (window)), a control condition (e.g., target object *hart*, distractor words *kano* (canoe), *zetel*, and *raam*), a named distractor condition (e.g., target object *raam*, distractor words *hart, harp*, and *zetel*), and a named distractor control condition (e.g., target object *raam*, distractor words *hart, kano,* and *zetel*).

Four stimulus lists were constructed (Appendix 1), each containing 48 displays. In two lists, competitors in the experimental displays were written words that were phonologically related to the names of the target objects. The phonological competitors were overlapping with the target names on 1-3 word-initial phonemes (e.g., 'broek' (/bruk/ - 'broer' (/brur/; *trousers* - *brother*). The same two lists implemented the named distractor conditions for the semantically related items. In two further lists, competitors in the experimental displays were written words that were category coordinates of the target object names. The category coordinates were selected by consulting the updated version of the Battig and Montague (1969) category norms (Van Overschelde, Rawson, & Dunlosky, 2004). We used the conceptual categories of body part, carpenter's tool, fruit, kitchen utensil, musical instrument, building part, thing that flies, thing that women wear, type of clothing, type of reading material, and vegetable. These two lists implemented the named distractor conditions for the phonologically related items.

The four lists were counterbalanced, so that within each list, 12 displays occurred in the experimental condition, 12 occurred in the control condition, 12 occurred in the named distractor condition, and 12 occurred in the named distractor control condition. Across the lists, one version of each display occurred once. Each list was presented to 12 participants. Table 1 summarizes the design.

Insert Table 1 about here

Procedure

Participants were seated at a comfortable distance, with their eyes approximately 50 cm from the display, in front of a 17 in. display. Eye movements were recorded using an EyeLink 1000 eye-tracker. We recorded speech to .WAV files using an ASIO driver, and manually measured naming latencies using a speech waveform editor.

We instructed participants to name the visual object in the display. They were asked to fixate a central fixation cross which appeared two seconds prior to the onset of the visual display. Each trial was terminated by the experimenter after an object had been named. The experiment lasted approximately 30 minutes.

Results

We only included responses that matched the expected response exactly, and so excluded trials in which participants failed to respond or named the object with a different name (e.g., morphological variant, synonym, subordinate, superordinate, visual error, semantically similar word) (21.19% overall). In none of these errors did the participants name a distractor word instead of the picture.

Naming latencies and error frequencies

In the condition with a phonological competitor word in the display and its control condition there were three items with more than 40% errors (*celery, flute, and screw*). These items were therefore discarded from analysis. From the remaining data set, trials with errors were excluded (5.8%). Trials with reaction times larger than 4000 ms (1.2%) and with reaction times that differed more than 3 standard deviations from the condition means (2.8%) were considered outliers and excluded. The mean naming latency in the condition with a phonological competitor word (1162 ms; SE = 38) did not differ from that in the control condition with all unrelated distractors (1168 ms; SE = 43), $t_1 < 1$ and $t_2 < 1$. The number of errors in the condition with a phonological competitor (15) did not differ from that in the corresponding control condition (14), $t_1 < 1$ and $t_2 < 1$.[1]

In the condition with a category coordinate competitor in the display and its control condition, there were five items with more than 40% errors (*teepee, pick, neck, orange,* and *nail*). These items were discarded from analysis. From the remaining data set, trials with errors were excluded (8.3%), and so were trials with naming latencies longer than 4000 ms (0.4%) and trials that differed more than three standard deviations from the condition means (2.2%). The mean naming latency in the condition with a category coordinate competitor (1396 ms; SE = 50) did not differ from that in the corresponding control condition (1339 ms; SE = 49), $t_1(23) = 1.62$, p > .1; t_2 (18) = 1.24, p > .1. The number of errors in the condition with a category coordinate competitor (18) did not differ from that in the corresponding control condition (20), $t_1 < 1$ and $t_2 < 1$.[2]

Analysis of eye-movements

Figure 2 shows a time-course graph that illustrates the fixation proportions at 20 ms intervals to the various types of stimuli over the course of the average trial. In computing these values, eye position was categorized according to the currently fixated quadrant. Proportion of fixations to the distractors was averaged across the unrelated distractor words. Zero represents the onset of the visual display.

Insert Figure 2a and 2b about here

The graphs show that as time unfolds fixation proportions to the target diverged from fixation proportions to the competitors and unrelated distractors for both types of experimental manipulations. In the condition with the category coordinate in the display (Figure 2b), fixation proportions to competitor and unrelated distractors never diverge. Critically, in the condition with the phonological competitor in the display (Figure 2a), there are more fixations to the competitor than the unrelated distractors shortly after speech onset.

In order to establish whether increased fixations occurred during the (estimated) time intervals when speech perception has access to inner and overt speech, we conducted separate statistical analyses for four intervals of 150 ms, corresponding respectively to (i) first possible inner speech-mediated eye-movements given a more liberal estimate of the head start of monitoring on voice onset; (ii) first possible inner speech-mediated eye-movements given a more conservative estimate of this head start; (iii) first region of 150 ms, given overt-speech mediated eye-movements; (iv) second region of 150 ms, given overt-speech mediated eye-movements. We will only report the analysis of the phonological conditions, but we note that in each of these intervals, there were significantly more fixations to the target than to category coordinate competitor words and to the unrelated distractor words (all $p < .001$), but fixation proportions to category coordinate competitors and unrelated distractors never reliably diverged (see Figure 2b).

(i) Fixation proportions from -50 ms to 100 ms after speech onset. Given Levelt's (1989) estimate of a 250 ms head start of monitoring on speech onset, and allowing 200 ms for programming a linguistically-mediated eye-movement, one would expect to observe the first inner-speech mediated eye-movents in this time interval. Table 2a shows the fixation proportions to the different stimuli during this time interval. Participants fixated the target

objects significantly more than the written phonological competitor words ($t_1(23) = 11.02$, $p < .001$; $t_2(20) = 8.98$, $p < .001$) and the written unrelated distractors ($t_1(23) = 11.41$, $p < .001$; $t_2(20) = 8.80$, $p < .001$). There was no difference between the written phonological competitor words and the written unrelated distractors ($t_1(23) = 1.03$, $p > .1$; $t_2(20) = 1.82$, $p > .1$).

Insert Table 2 about here

(ii) Fixation proportions from 55 ms to 205 ms after speech onset. Given Indefrey and Levelt's (2004) estimate of a 145 ms head start of monitoring on speech onset, and allowing 200 ms for programming a linguistically-mediated eye-movement, the first inner-speech mediated eye-movents should be observed in this time interval. Table 2b shows the fixation proportions to the different stimuli during this time interval. Participants fixated the target objects significantly more than the written phonological competitor words ($t_1(23) = 12.18$ $p < .001$; $t_2(20) = 9.34$, $p < .001$) and the written unrelated distractors ($t_1(23) = 11.01$, $p < .001$; $t_2 (20) = 8.47$, $p < .001$). There was no difference between the written phonological competitor words and the written unrelated distractors ($t_1 < 1$; $t_2 (20) = 1.06$, $p > .1$).

(iii) Fixation proportions from 200 ms to 350 ms after speech onset. Allowing for 200 ms to program a linguistically mediated eyemovement, this is the earliest interval in which one can expect to observe eye-movements mediated by overt speech perception.Table 2c shows the fixation proportions to the different stimuli during this time interval. Participants fixated the target objects significantly more than the written phonological competitor words ($t_1(23) = 10.29$, $p < .001$; $t_2(20) = 8.78$, $p < .001$) and the written unrelated

distractors ($t_1(23) = 10.02$, $p < .001$; $t_2(20) = 8.33$, $p < .001$). There was no difference between the written phonological competitor words and the written unrelated distractors ($t_1 < 1$ and $t_2 < 1$).

(iv) Fixation proportions from 350 ms to 500 ms after speech onset. This interval directly follows the earliest interval during which one can expect to observe eye-movements mediated by overt speech perception. Table 2d shows the fixation proportions to the different stimuli during this time region. Participants fixated the target objects significantly more than the written phonological competitor words ($t_1(23) = 6.80$, $p < .001$; $t_2(20) = 6.94$, $p < .001$) and the written unrelated distractors ($t_1(23) = 9.22$, $p < .001$; t_2 $(20) = 9.11$, $p < .001$). Critically, the participants fixated the phonological competitor words significantly more than the written unrelated distractors ($t_1(23) = 3.10$, $p < .01$; $t_2(20) = 2.09$, $p < .05$). In the control condition, in which an unrelated object (the 'named distractor') had to be named, there were no differences between the written phonological competitor words and the written unrelated distractors ($t_1 < 1$ and $t_2 < 1$).

Note that the phonological effect in a similar experiment investigating other-perception (Huettig & McQueen, 2007, Experiment 4) became first significant in the 300 ms to 400 ms time window[3] (and more reliably so in the 400 ms to 500 ms window) but not before. Thus, the timing of the attentional shift in the present production experiment was similar to the one in Huettig and McQueen's perception experiment.

Discussion

Our experiment shows that listening to one's own speech drives eye-movements in the visual world in a very similar way as listening to someone else's speech (Huettig & McQueen, 2007). In both modalities, shortly after the acoustic onset of a word, participants fixate phonologically-related written words more often than unrelated words. Importantly, the time-course of fixating phonological competitors observed here in self-perception was very similar to that in other-perception (Huettig & McQueen, 2007), even though perceptually-based monitoring theories predict that phonological information is available much earlier (145 - 250 ms) in self-perception than other-perception. It thus appears that the internal monitoring channel does not rely on speech perception.

In addition to our critical manipulation of phonological overlap, we included several control conditions. The control condition with category coordinate competitor words, showed that participants were *not* more likely to fixate the competitors than the unrelated distractors. Thus the data pattern (phonological but not semantic competition when written word displays are used) is comparable to that observed in Huettig and McQueen (2007). This further supports the hypothesis that listening to one's own speech is similar to listening to someone else's speech.

As a further control, we compared naming latencies with or without the presence of a competitor. Presence of a competitor did not affect the speed or accuracy of picture naming, suggesting that the written competitor manipulation did not affect the word production process itself. We note that phonological and semantic context effects are typically found in the picture-word interference task (in which speakers name an object and ignore a written word distractor which is superimposed on that object). Presumably, such context effects did not occur presently, because there were multiple written distractor words and there was spatial displacement between picture and distractor word (the latter manipulation is known to greatly reduce Stroop-like effects, Risko, Stolz, & Besner, 2005).

As a final control, we included a condition in which the former target was replaced by its written name, and a former unrelated written distractor was replaced by its picture (the "named distractor condition"). This condition showed that it is *not* the case that the competitor items intrinsically attracted more visual attention than the unrelated distractors; it was also *not* the case that having two similar items in a display (e.g., the written words *heart* and *harp*) intrinsically drew visual attention to these similar items.

Before we discuss the theoretical implications of our findings, we need to address two potential caveats. First, the reaction time analysis showed that naming latencies here were on average much longer (> 1000 ms) than those in picture-word interference (about 600 ms; Indefrey & Levelt, 2004). One might therefore ask whether production under the current task conditions is representative of spontaneous speech production. One reason for the longer naming latencies may be that the participants needed to select the target picture out of four stimuli, whereas in picture-word interference only two stimuli are presented. Additionally, picture-word interference experiments usually include a training phase in which participants learn to use a particular name with a particular picture. Finally, such experiments usually present each picture multiple times, so that the naming latencies can benefit from repetition priming. In contrast, we did not train our participants and presented each picture only once, a procedure we believe more closely approximates the naming of objects in a scene under naturalistic circumstances. In fact, our latencies are similar to those in other studies that tested picture naming under these conditions (e.g., 1090 ms in Severens et al.'s 2005 Dutch norms; see also Bates et al., 2003 for similar results in seven further languages, including English).

A related concern is whether it is legitimate to use Indefrey and Levelt's (2004) estimate of the head start of inner speech perception on production; after all, Indefrey and Levelt's fractioning of naming latencies was based on results from picture-word interference. We assume however that the procedural differences have the largest consequences for the earliest stages of picture naming, in particular object recognition and perhaps lexical selection (as is suggested by our categorization of naming errors, footnotes 1 and 2). It is possible of course that phonological encoding and articulatory planning also proceed more slowly in the present experiment as compared to picture-word interference. If so, this would not challenge our conclusions: in fact, on that story phonological encoding begins even earlier relative to the onset of overt speech, so that eye-movements directed by inner speech should have occurred even earlier than on the estimates we have used. But the data showed no such early eye-movements to the distractors.

A second potential criticism is that internal monitoring does in fact engage speech perception, but without the behavioral consequences that speech perception usually has in visual world paradigms, namely increased number of looks to phonologically related items. On one such account, participants may have preferred to look at the target object (e.g., *hart*, [heart]) because its name constitutes an even better match with internal speech than the phonologically related distractor word (*harp)* does. Note however that the assumption of perceptual monitoring theories is that in internal monitoring the speech perception system inspects a phonetic (or phonological) code as it unfolds in time, analogous to listening to overt speech. A great number of studies with other-produced overt speech have clearly indicated that eye-movements are driven to objects with an initial phonological match (e.g., candy), even when an object with a full match (candle) is present.

Of course, a difference between listening to others and listening to oneself is that only in the latter case, the listener has always prior knowledge of the speaker's intention (Hartsuiker, 2006*).* It is conceivable that this knowledge (e.g., that the next word should be *heart)* constrains perception, so that a distractor word not fully matching the intention (*harp)* is never considered as a candidate and hence never attended. However, our data clearly indicate that attention was driven by one's overt speech in a similar way as by someone else's speech, which makes it unlikely that knowledge of the intention influenced perception of self-produced overt speech. The only way in which this account may be maintained, is by making the strong assumption that perception of inner speech is more strongly guided by top-down information (i.e., by the intention) than perception of overt speech.

Another such account has it that the target objects captured and held visual attention during the entire naming process, preventing any early looks to the competitor words. The present study as well as other studies (Griffin, 2004; Meyer, 2004) indeed showed that people look at the objects they name. However, several studies by Meyer and colleagues (e.g., Meyer, 2004) have demonstrated that participants usually shift their gaze to another object well before articulation onset of the target object's name (i.e., 200 ms - 300 ms). Indeed there is no plausible reason why participants could disengage their attention from the object during the first 100 ms that acoustic information from the overt speech became available to the eye movement system (as our study shows) but should not have been able to do so 200 - 300 ms earlier. Thus we can safely assume that during the time window of interest to us (-50 ms to 200 ms after articulation) participants can disengage their attention from the picture.

A final version of such an account is that eye-movements in reaction to speech, are driven by a lower-level representation of speech (e.g., a sub-phonemic code) than is generated by the internal monitoring channel (e.g., a phonemic code). If that were the case, internal-monitoring would bypass the system that drives eye-movements, and so no eye-movements in response to the internal channel can be expected. Note that some perceptually-based monitoring theories claim that the internal monitor inspects a phonetic code (Levelt, 1989) whereas others claim this code is phonological (Wheeldon & Levelt, 1995; but see Vigliocco & Hartsuiker, 2002, for criticism of this claim). Similarly, based on the pattern of speech errors in overt and silent speech, Oppenheim and Dell (2008) recently concluded that silent speech is impoverished at the subphonemic level but is intact at the phonemic level (although they acknowledged the possibility that their conclusion may not hold for inner speech preceding overt articulation).

However, even granting the possibility that inner speech is impoverished, we still have little reason to believe that our results could be explained by a “bypass” account, which would entail that eye-movements here would be driven by sub-phonemic rather than phonemic matches between speech and competitor items. This is because our phonological competitor items were written words. Studies in visual word recognition demonstrate that written words are recoded phonologically (see Frost, 1998, for review), and so the representations of the written phonological competitor word and the spoken word match at a phonological level. Theories of visual word recognition typically assume that written words are recoded into a phonemic code (e.g., via Grapheme-Phoneme Conversion rules) or a supra-phonemic code (whole-word phonology) but not into a subphonemic code[4](see Huettig & McQueen, 2007, for further discussion on the mapping processes between spoken words and visual stimuli).

In sum, we reject several alternative accounts according to which our production task may be more unrepresentative of natural speech production than other paradigms, or according to which internal speech monitoring engages speech perception, but without this form of perception driving eye-movements. Our data, in conjunction with the neuropsychological studies cited in the introduction, are more compatible with a view on which internal monitoring does not engage speech perception.

Our conclusion differs from the only study that has provided evidence for a perceptually-based internal channel (Özdemir et al., 2007). It is important to note, however, that their findings were based on a silent speech task (*does the name of this picture contain a /b/?)* which (by definition) does not involve overt production. This situation is rather different from self-monitoring of naturalistic speech, where the production of an internal speech representation is followed shortly by overt articulation (Oppenheim & Dell, 2008). This is one reason why Vigliocco and Hartsuiker (2002) argued against comprehension-based internal monitoring, because the presence of both inner and overt speech (two identical phonological codes that are perceived with only a slight temporal asynchrony) would create considerable problems for the speech perception system, such as interference between the codes and the subjective experience of "echoes". They therefore proposed a production-based internal channel, while acknowledging the possibility that speakers can listen to inner speech when producing speech *silently*. On this account it is not surprising that there are perceptual effects in silent phoneme monitoring.

Interestingly, there were no increased looks to semantically related distractors; note that this was also true in Huettig and McQueen's (2007, Experiment 4) perception experiment. This is further evidence that with printed word displays, participants focus on the possibility of phonological matches because the display consists of easily accessible orthographic renditions of the sound forms of words (Huettig & McQueen, 2007; McQueen & Viebahn, 2007). Our finding that we pay close attention to phonology when listening to ourselves has a further important implication. It suggests that we can immediately monitor our overt speech at the level of phonology (Slevc & Ferreira, 2006) rather than wait until speech comprehension has cascaded further to make a comparison at the level of "parsed speech", which would be a representation of the input in terms of its phonological, but also morphological, syntactic, and semantic composition (Levelt, 1989).

The present results thus offer no support for a perceptual theory of inner speech monitoring. We have to acknowledge that there are presently no elaborated theories of the alternative viewpoint, namely production-internal monitoring. This latter viewpoint is sometimes criticized (e.g., by Levelt, 1989) because it would require reduplication of information processing; that is, it would be baroque if one and the same system both computed a particular response (say the word “cat” on the basis of its concept) and, for verification purposes, another version of that same response. We completely agree that a theory which involves reduplication is not parsimonious. However, it is a theoretical possibility that a production monitor could work without reduplication. As mentioned in the introduction, a production monitor might inspect the global dynamics at each level of processing (say the lexical level). If all is going to plan, one would expect one representation (i.e., the correct lexical node) to be much more active than all other representations; but if an error is (about to be) made, one would expect at least two units to be highly active: one corresponding to the correct word, and one corresponding to the error. It remains to be seen what the explanatory power of this proposal is with respect to existing findings in the monitoring literature. One issue, for example, is whether conflict monitoring at a lexical layer might account for the lexical bias effect (the finding that phonological errors result more often in real words as opposed to non-words than chance would predict), which is often ascribed to self-monitoring (see Hartsuiker, 2006, for review) and for modulations of this effect as a function of context (Hartsuiker, Corley, & Martensen, 2005). Although further research is surely needed to make this proposal more explicit and test it empirically, it is clear that such an account predicts no perceptual effects of monitoring inner speech: on this account, monitoring of planned speech is the business of production-internal mechanisms only.

In conclusion, speakers listen to their own external, but not internal, speech in speech production. The phonological code in perception has immediate consequences for eye-movement behavior with printed word displays.

Acknowledgements

This research was sponsored by a grant from FWO, the Fund for Research - Flanders (G.0181.05). We thank Evelyne Lagrou and Ilse Tydgat for their assistence. Part of this research was presented at the AMLaP 2008 conference in Cambridge, England.

References

Altmann, G.T.M., & Kamide, Y. (2004). Now you see it, now you don't: Mediating the mapping between language and visual world. In J. Henderson & F. Ferreira (Eds.) *The interface of language, vision, and action: Eye movements and the visual world* (pp. 347-386). Hove, England: Psychology Press.

Battig, W.F. & Montague, W.E. (1969). Category norms for verbal items in 56 categories: A replication and extension of the Connecticut category norms. *Journal of Experimental Psychology Monograph*, *80*, 1-46.

Bates, E., D'Amico, S., Jacobsen, T., Székely, A., Andonova, E., Devescovi, A., Herron, D., Lu, C. C., Pechmann, T., Pléh, C., Wicha, N., Federmeier, K., Gerdjikova, I., Gutierrez, G., Hung, D., Hsu, J., Iyer, G., Kohnert, K., Mehotcheva, T., Orozco-Figueroa, A., Tzeng, A., & Tzeng, O. (2003). Timed picture naming in seven languages. *Psychonomic Bulletin & Review*, *10*, 344-380.

Botvinick, M.M., Braver, T.S., Barch, D.M., Carter, C.S., & Cohen, J.D. (2001). Conflict monitoring and cognitive control. *Psychological Review, 108*, 624–652.

Cooper, R.M. (1974). The control of eye fixation by the meaning of spoken language: A new methodology for the real-time investigation of speech perception, memory, and language processing. *Cognitive Psychology, 6*, 84-107.

Dell, G.S. & Repka R.J. (1992). Errors in inner speech. In Baars B.J. (Ed.). *Experimental slips and human error: exploring the architecture of volition* (pp. 237–262). New York: Plenum Press.

Frost, R. (1998). Towards a strong phonological theory of visual word recognition: True issues and false trails. *Psychological Bulletin, 123, 71-99.*

Griffin, Z.M. (2004). Why look? Reasons for speech-related eye movements. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action* (pp. 213-248). Hove, England: Psychology Press.

Hartsuiker, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language and Cognitive Processes, 21,* 856-891.

Hartsuiker, R. J., Corley, M., & Martensen, H. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related Beply to Baars, Motley, and MacKay (1975). *Journal of Memory and Language*, *52*, 58 - 70.

Hartsuiker, R.J. & Kolk, H.H.J. (2001). Error monitoring in speech production: A computational test of the perceptual loop theory. *Cognitive Psychology, 42*, 113–157.

Huettig, F., & Hartsuiker, R.J. (2008). When you name the pizza you look at the coin and the bread: Eye movements reveal semantic activation during word production. *Memory & Cognition, 36,* 341-360.

Huettig, F., & McQueen, J.M. (2007). The tug of war between phonological, semantic, and shape information in language-mediated visual search. *Journal of Memory and Language, 54*, 460-482.

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

Lackner, J. R., & Tuller, B.H. (1979). Roles of efference monitoring in the detection of self-produced speech errors. In W.E. Cooper & E.C.T. Walker

(Eds.),*Sentence processing. Psycholinguistic studies presented to Merrill Garrett* (pp. 281–294). Hillsdale, NJ: Erlbaum.

Laver, J. (1980). Monitoring systems in the neurolinguistic control of speech production. In V. A. Fromkin (Ed.), *Errors in linguistic performance: slips of the tongue, ear, pen, and hand.* New York, Academic Press.

Levelt W.J.M. (1983). Monitoring and self-repair in speech. *Cognition*, *14***,** 41–104.

Levelt, W. J. M. (1989). *Speaking. From intention to articulation*. Cambridge, MA: MIT Press.

Lukatela, G., Eaton, T., Lee, C., & Turvey, M. T. (2001). Does visual word identification involve a sub-phonemic level? *Cognition, 78,* B41-B52*.*

Mattson, M., & Baars, B.J. (1992). Error-minimizing mechanisms: Boosting or editing? In B. J. Baars (Ed.), *Experimental Slips and Human Error: Exploring the Architecture of Volition* (pp. 263-287). New York: Plenum.

Marshall R.C., Rappaport B.Z., Garcia-Bunuel L. (1985). Self-monitoring behavior in a case of severe auditory agnosia with aphasia. *Brain and Language*, *24*, 297–313.

Marshall J., Robson J., Pring T., & Chiat S. (1998). Why does monitoring fail in jargon aphasia: Comprehension, judgment and therapy evidence. *Brain and Language*, *63*, 79-107.

Marslen-Wilson, W.D. (1990). Activation competition and frequency in lexical access. In G.T.M. Altmann (Ed.), *Cognitive models of speech processing: psycholinguistic and computational perspectives* (pp. 148-172). Cambridge, MA: MIT Press.

McNamara, P., Obler, L., Au, R., Durso, R., & Albert, M. (1992). Speech monitoring skills in Alzheimer's Disease, Parkinson's disease and normal aging. *Brain & Language, 42*, 38-51.

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

Meyer, A.S. (2004). The use of eye tracking in studies of sentence generation. In J. M. Henderson & F. Ferreira (Eds.), *The interface of language, vision and action* (pp. 191-212). Hove, England: Psychology Press.

Motley, M.T., Camden, C.T., & Baars, B.J. (1982). Covert formulation and editing of anomalies in speech production: evidence from experimentally elicited slips of the tongue. *Journal of Verbal Learning and Verbal Behavior, 21,* 578–594.

Nickels, L.A., & Howard, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, *31*, 209-237.

Oomen, C.C.E., Postma, A., & Kolk, H.H.J. (2005). Speech monitoring in aphasia: Error detection and repair behaviour in a patient with Broca's aphasia. In R.J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.) *Phonological Encoding and Monitoring in Normal and Pathological Speech* (pp. 209 - 225). Hove, England: Psychology Press.

Oppenheim, G. M., & Dell, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition, 106*, 528-537.

Özdemir, R., Roelofs, A., & Levelt, W. J. M. (2007). Perceptual uniqueness point effects in monitoring internal speech, *Cognition, 105*, 457-465.

Postma, A. (2000). Detection of errors during speech production. A review of speech monitoring models. *Cognition*, *77*, 97–131.

Risko, E. F., Stolz, J. A., & Besner, D. (2005). Basic processes in reading: Is visual word recognition obligatory? *Psychonomic Bulletin & Review, 12*, 119-124.

Schriefers, H., Meyer, A. S., & Levelt, W. J. M. (1990). Exploring the time-course of lexical access in language production: Picture-word interference studies. *Journal of Memory and Language, 29*, 86-102.

Severens, E., Van Lommel, S., Ratinckx, E., & Hartsuiker, R.J. (2005). Timed picture naming norms for 590 pictures in Dutch. *Acta Psychologica*, *119*, 159-187.

Slevc, L.R. & Ferreira, V.S. (2006). Halting in single-word production: A test of the perceptual-loop theory of speech monitoring. *Journal of Memory and Language, 54*, 515-540.

Tanenhaus, M.K., Spivey-Knowlton, M.J., Eberhard, K.M., & Sedivy, J.C. (1995). Integration of visual and linguistic information in spoken language comprehension. *Science*, *268*, 1632-1634.

Van Overschelde, J.P., Rawson, K. A., & Dunlosky, J. (2004). Category norms: An updated and expanded version of the Battig and Montague (1969) norms. *Journal of Memory and Language, 50*, 289-335.

Vigliocco, G. & Hartsuiker, R.J. (2002). The interplay of meaning, sound, and syntax in language production. *Psychological Bulletin*, *128*, 442-472.

Wheeldon, L.R. & Levelt, W.J.M. (1995). Monitoring the time course of phonological encoding. J*ournal of Memory and Language*, *34*, 311-334.

Footnotes

[1] For archival purposes, we report the distribution of error types. The 29 naming errors consisted of 8 morphological variants (e.g., *vlieger => "windvlieger", kite => "wind kite";* windvlieger is a morphologically transparent neologism), 9 synonyms (e.g., *bot => been*; both mean *bone*), 6 semantically similar names (e.g., *rok (skirt) => jurk (dress)),* 3 names at a different description level than intended, such as subordinates, superordinates, parts, or larger wholes (e.g., *brief (letter)=> post (mail)*), and 3 visual errors (*aardappel (potato) => vlek (spot*)).

[2] For archival purposes, we report the distribution of error types. The 38 naming errors consisted of 11 morphological variants (*ketting (necklace) => halsketting (necklace; hals* means *neck*), 7 semantically similar words (*perzik (peach) => pruim (plumb*)), 11 synonyms (*touw (rope) => koord (cord)*), 2 visual errors (*tuinslang (garden hose) => touw (rope)),* and 7 names at a different description level (*bliksem (lightning) => onweer (thunderstorm*)).

[3] Huettig and McQueen (2007) analyzed separate 100 ms time windows in their study. We chose 150 ms here to be consistent with the 150 ms intervals of the possible inner speech-mediated eye-movements.

[4] As far as we are aware, only one study provided evidence that visual word recognition involves sub-phonemic information (Lukatela, Eaton, Lee, & Turvey, 2001). But note that these authors explicitly argue that such involvement "is not evidence against a segmental level" (p. 42).

Appendix 1

a) List A

item	target	competitor	unrelated distractor	*named* distractor	distractor 3
1	lamp *(lamp)**	lamb *(lam)*	hand	watermelon	
2	finger *(vinger)*	fin *(vin)*	house	cutting board	
3	heart *(hart)*	harp *(harp)*	couch	window	
4	nose *(neus)*	net *(net)*	mixer	submarine	
5	bone *(bot)*	bock *(bok)*	piano	tv	
6	ruler *(lat)*	Chinese latern *(lampion)*	airplane	cherry	
7	screw *(schroef)*	typewriter *(schrijfmachine)*	zebra	alarm clock	
8	apple *(appel)*	anchor *(anker)*	chair	bat	
9	banana *(banaan)*	battery *(batterij)*	ostrich	radish	
10	spoon *(lepel)*	liver *(lever)*	glove	paper	
11	flute *(dwarsfluit)*	dwarf *(dwerg)*	sailboat	fist	
12	saxophone *(saxofoon)*	salad *(salade)*	frog	mosquito	
13	door		shark	arm	lettuce
14	kite		spatula	dresser	tiger
15	lipstick		penguin	table	drum
16	skirt		peacock	dolphin	lizard
17	belt		pig	bee	doll
18	pants		volcano	bug	doctor
19	scarf		eagle	fly	sheep
20	letter		bird	bus	butterfly
21	corn		boot	bowl	turtle
22	tomato		dentist	worm	bat
23	potato		squirrel	gun	pineapple
24	celery		butcher	speaker	owl
25	orange		shoulder	elephant	motorcycle
26	peach		block	accordion	turkey
27	sun		slipper	goat	desk
28	hammer		dog	palm tree	blimp
29	pick		rose	mouse	trumpet
30	rake		baby	pumpkin	dinosaur
31	nail		lion	tractor	radio
32	ring		ear	cow	puzzle

33	necklace		spider	grapes	cook
34	rope		duck	skateboard	pot
35	water hose		bell	parrot	fireman
36	tie		teeth	yo-yo	leaf
37	knife	cup	mushroom	swan	
38	leg	molar	fox	onion	
39	thumb	eye	chicken	clamp	
40	toe	chest	cear	can opener	
41	neck	lips	harp	anvil	
42	bricks	roof	rhino	rocket	
43	bed	stool	banjo	axe	
44	mountain	tree	violin	hat	
45	teepee	trailer	donkey	lawnmower	
46	lightning	rain	police man	paintbrush	
47	train	helicopter	organ	stairs	
48	plate	stove	horse	sled	

* in italics the Dutch translation for the items in the phonological overlap conditions

b) List B

item	target	competitor	unrelated distractor	*named* distractor	distractor 3
1	lamp		hand	watermelon	whale
2	finger		house	cutting board	bench
3	heart		couch	window	canoe
4	nose		mixer	submarine	church
5	bone		piano	tv	whistle
6	ruler		airplane	cherry	pelican
7	screw		zebra	alarm clock	kangaroo
8	apple		chair	bat	deer
9	banana		ostrich	radish	bicycle
10	spoon		glove	paper	car
11	flute		sailboat	fist	sock
12	saxophone		frog	mosquito	foot
13	door *(deur)*	good-for-nothing *(deugniet)*	shark	arm	
14	kite *(vlieger)*	elderbush *(vlierstruik)*	spatula	dresser	
15	lipstick *(lippenstift)*	lift *(lift)*	penguin	table	
16	skirt *(rok)*	seal *(rob)*	peacock	dolphin	
17	belt *(riem)*	reed *(riet)*	pig	bee	
18	pants *(broek)*	brother *(broer)*	volcano	bug	
19	scarf *(sjaal)*	stencilplate *(sjabloon)*	eagle	fly	
20	letter *(brief)*	bride *(bruid)*	bird	bus	
21	corn *(mais)*	folder *(map)*	boot	bowl	
22	tomato *(tomaat)*	tower *(toren)*	dentist	worm	
23	potato *(aardappel)*	ape *(aap)*	squirrel	gun	
24	celery *(selder)*	napkin *(servet)*	butcher	speaker	
25	orange	pear	shoulder	elephant	
26	peach	strawberry	block	accordion	
27	sun	cloud	slipper	goat	
28	hammer	drill	dog	palm tree	
29	pick	watering can	rose	mouse	
30	rake	wheelbarrow	baby	pumpkin	
31	nail	saw	lion	tractor	
32	ring	dress	ear	cow	
33	necklace	heel	spider	grapes	

34	rope	gun	duck	skateboard	
35	water hose	shovel	bell	parrot	
36	tie	shoe	teeth	yo-yo	
37	knife		mushroom	swan	jacket
38	leg		fox	onion	telephone
39	thumb		chicken	clamp	beaver
40	toe		cear	can opener	sweater
41	neck		harp	anvil	monkey
42	bricks		rhino	rocket	shirt
43	bed		banjo	axe	giraffe
44	mountain		violin	hat	rabbit
45	teepee		donkey	lawnmower	alligator
46	lightning		police man	paintbrush	truck
47	train		organ	stairs	wolf
48	plate		horse	sled	ant

c) List C

item	target	competitor	unrelated distractor	*named* distractor	distractor 3
1	lamp *(lamp)*	lamb *(lam)*	hand	watermelon	
2	finger *(vinger)*	fin *(vin)*	house	cutting board	
3	heart *(hart)*	harp *(harp)*	couch	window	
4	nose *(neus)*	net *(net)*	mixer	submarine	
5	bone *(bot)*	bock *(bok)*	piano	tv	
6	ruler *(lat)*	Chinese latern *(lampion)*	airplane	cherry	
7	screw *(schroef)*	typewriter *(schrijfmachine)*	zebra	alarm clock	
8	apple *(appel)*	anchor *(anker)*	chair	bat	
9	banana *(banaan)*	battery *(batterij)*	ostrich	radish	
10	spoon *(lepel)*	liver *(lever)*	glove	paper	
11	flute *(dwarsfluit)*	dwarf *(dwerg)*	sailboat	fist	
12	saxophone *(saxofoon)*	salad *(salade)*	frog	mosquito	
13	door		shark	arm	lettuce
14	kite		spatula	dresser	tiger
15	lipstick		penguin	table	drum
16	skirt		peacock	dolphin	lizard
17	belt		pig	bee	doll
18	pants		volcano	bug	doctor
19	scarf		eagle	fly	sheep
20	letter		bird	bus	butterfly
21	corn		boot	bowl	turtle
22	tomato		dentist	worm	bat
23	potato		squirrel	gun	pineapple
24	celery		butcher	speaker	owl
25	orange		shoulder	elephant	motorcycle
26	peach		block	accordion	turkey
27	sun		slipper	goat	desk
28	hammer		dog	palm tree	blimp
29	pick		rose	mouse	trumpet
30	rake		baby	pumpkin	dinosaur
31	nail		lion	tractor	radio
32	ring		ear	cow	puzzle

33	necklace		spider	grapes	cook
34	rope		duck	skateboard	pot
35	water hose		bell	parrot	fireman
36	tie		teeth	yo-yo	leaf
37	knife	cup	mushroom	swan	
38	leg	molar	fox	onion	
39	thumb	eye	chicken	clamp	
40	toe	chest	cear	can opener	
41	neck	lips	harp	anvil	
42	bricks	roof	rhino	rocket	
43	bed	stool	banjo	axe	
44	mountain	tree	violin	hat	
45	teepee	trailer	donkey	lawnmower	
46	lightning	rain	police man	paintbrush	
47	train	helicopter	organ	stairs	
48	plate	stove	horse	sled	

d) List D

item	target	competitor	unrelated distractor	*named* distractor	distractor 3
1	lamp		hand	watermelon	whale
2	finger		house	cutting board	bench
3	heart		couch	window	canoe
4	nose		mixer	submarine	church
5	bone		piano	tv	whistle
6	ruler		airplane	cherry	pelican
7	screw		zebra	alarm clock	kangaroo
8	apple		chair	bat	deer
9	banana		ostrich	radish	bicycle
10	spoon		glove	paper	car
11	flute		sailboat	fist	sock
12	saxophone		frog	mosquito	foot
13	door *(deur)*	good-for-nothing *(deugniet)*	shark	arm	
14	kite *(vlieger)*	elderbush *(vlierstruik)*	spatula	dresser	
15	lipstick *(lippenstift)*	lift *(lift)*	penguin	table	
16	skirt *(rok)*	seal *(rob)*	peacock	dolphin	
17	belt *(riem)*	reed *(riet)*	pig	bee	
18	pants *(broek)*	brother *(broer)*	volcano	bug	
19	scarf *(sjaal)*	stencilplate *(sjabloon)*	eagle	fly	
20	letter *(brief)*	bride *(bruid)*	bird	bus	
21	corn *(mais)*	folder *(map)*	boot	bowl	
22	tomato *(tomaat)*	tower *(toren)*	dentist	worm	
23	potato *(aardappel)*	ape *(aap)*	squirrel	gun	
24	celery *(selder)*	napkin *(servet)*	butcher	speaker	
25	orange	pear	shoulder	elephant	
26	peach	strawberry	block	accordion	
27	sun	cloud	slipper	goat	
28	hammer	drill	dog	palm tree	
29	pick	watering can	rose	mouse	
30	rake	wheelbarrow	baby	pumpkin	
31	nail	saw	lion	tractor	
32	ring	dress	ear	cow	
33	necklace	heel	spider	grapes	

34	rope	gun	duck	skateboard	
35	water hose	shovel	bell	parrot	
36	tie	shoe	teeth	yo-yo	
37	knife		mushroom	swan	jacket
38	leg		fox	onion	telephone
39	thumb		chicken	clamp	beaver
40	toe		cear	can opener	sweater
41	neck		harp	anvil	monkey
42	bricks		rhino	rocket	shirt
43	bed		banjo	axe	giraffe
44	mountain		violin	hat	rabbit
45	teepee		donkey	lawnmower	alligator
46	lightning		police man	paintbrush	truck
47	train		organ	stairs	wolf
48	plate		horse	sled	ant

Table 1. Design of the experiment

Set	Rel	List			
		A	B	C	D
1	Phon	ND	ND contr	Exp	Contr
2	Phon	ND contr	ND	Contr	Exp
3	Sem	Contr	Exp	ND contr	ND
4	Sem	Exp	Contr	ND	ND contr

Note. Set = set of items (n = 12 per set). Rel = type of relatedness (Phon = phonological, Sem = semantic). List = Stimulus list. ND = Named distractor condition. ND contr = Named distractor control condition. Exp = Experimental condition. Contr = Control condition.

Table 2a

Fixation proportions (in %) during interval of 1st inner speech estimate

Condition	Phonological			Category		
Type of Picture	Target	Competitor	Distractor	Target	Competitor	Distractor
Fixation Proportion	57	8	7	55	6	4

Table 2b

Fixation proportions (in %) during interval of 2nd inner speech estimate

Condition	Phonological			Category		
Type of Picture	Target	Competitor	Distractor	Target	Competitor	Distractor
Fixation Proportion	56	8	8	54	7	4

Table 2c

Fixation proportions (in %) during 1st interval for external speech-mediated eye-movements

Condition	Phonological			Category		
Type of Picture	Target	Competitor	Distractor	Target	Competitor	Distractor
Fixation Proportion	53	9	8	54	6	4

Table 2d

Fixation proportions (in %) during 2nd interval for external speech-mediated eye-movements

Condition	Phonological			Category		
Type of Picture	Target	Competitor	Distractor	Target	Competitor	Distractor
Fixation Proportion	49	14	7	54	4	5

Figure Captions

Figure 1 a). An example of the type of visual displays used - experimental condition: target (e.g. hart - *heart*) , phonological competitor (harp - *harp*), and two unrelated distractors (zetel - *couch,* raam - *window*); b) control condition: target (e.g. hart - *heart*) and three unrelated distractors (kano - *canoe*, zetel - *couch,* raam - *window*); c) named distractor condition (e.g., raam - *window* is to be named, and occurs with harp - *harp,* hart - *heart*, and zetel - *couch*); d) named distractor control condition: all items were unrelated(e.g., raam - *window* is to be named, and occurs with kano - *canoe,* hart - *heart*, and zetel - *couch*).

Figure 2. Time-course graph showing the fixation probabilities to the target, competitors, and distractors in a) phonological condition (competitor present in the display) and b) category condition (competitor present in the display). Zero on the time axis refers to the display onset.

Figure 1

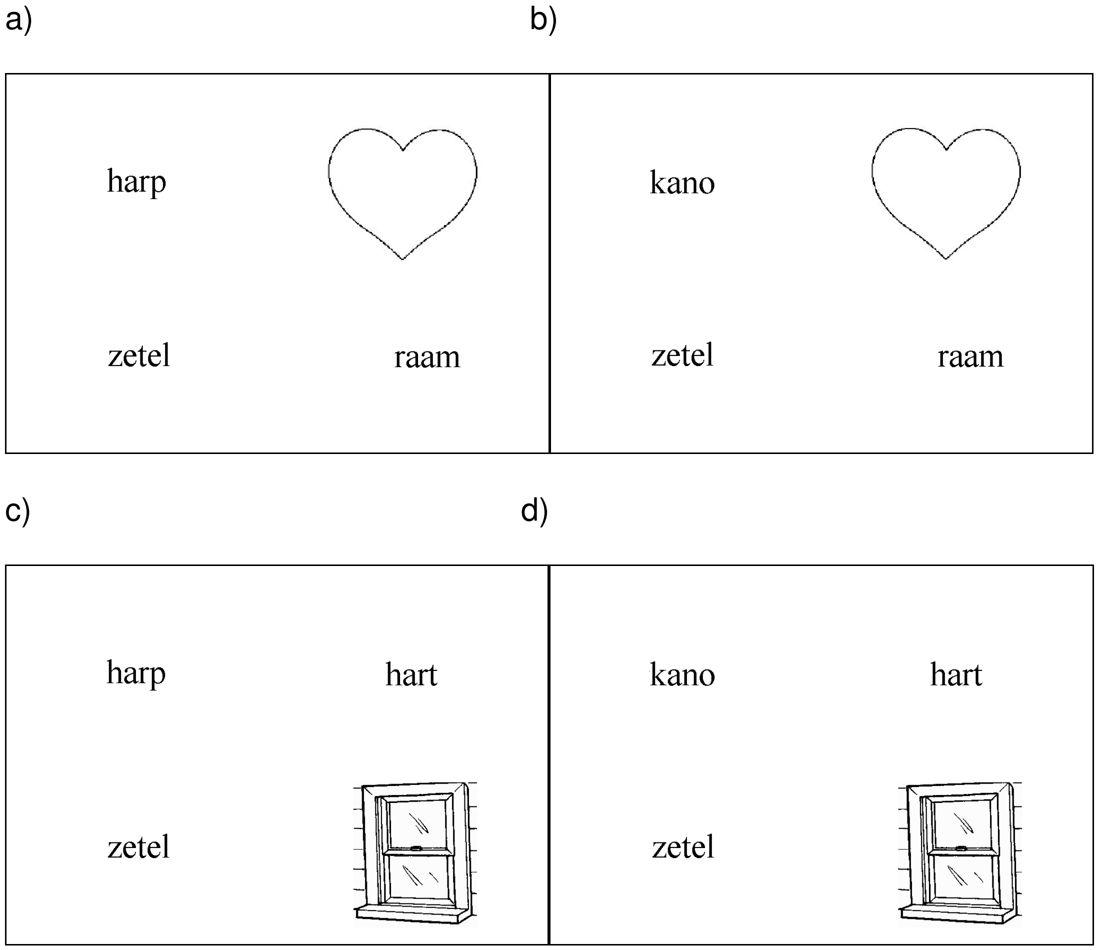

a)
b)
harp
zetel
raam
kano
zetel
raam
c)
d)
harp
hart
zetel
kano
hart
zetel

Figure 2a

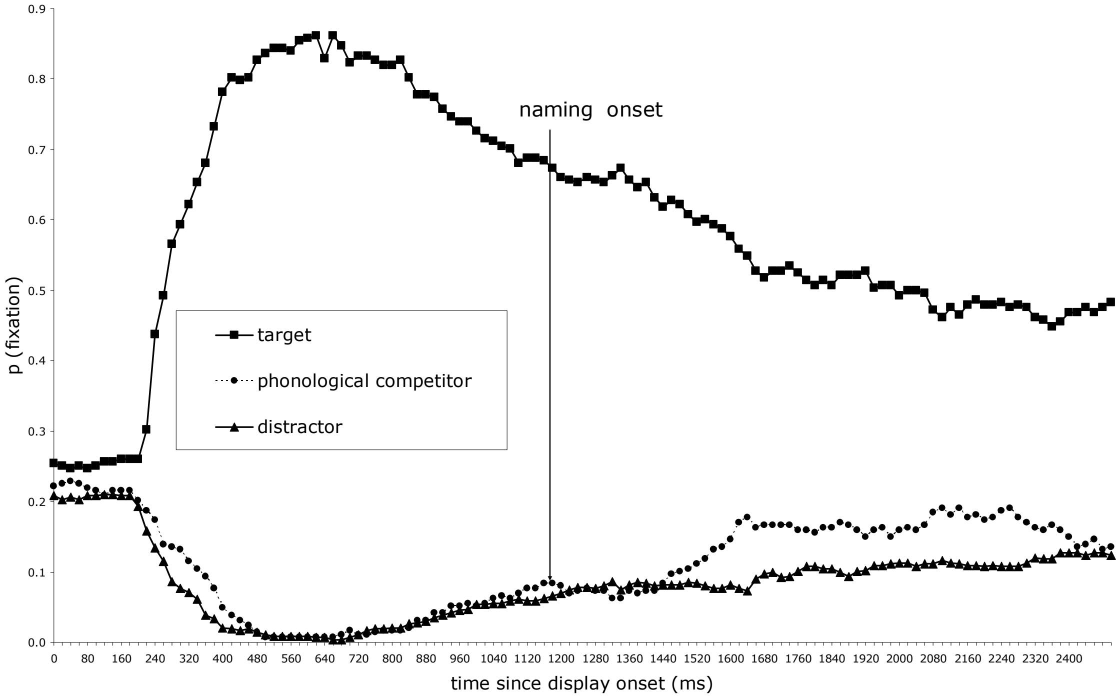
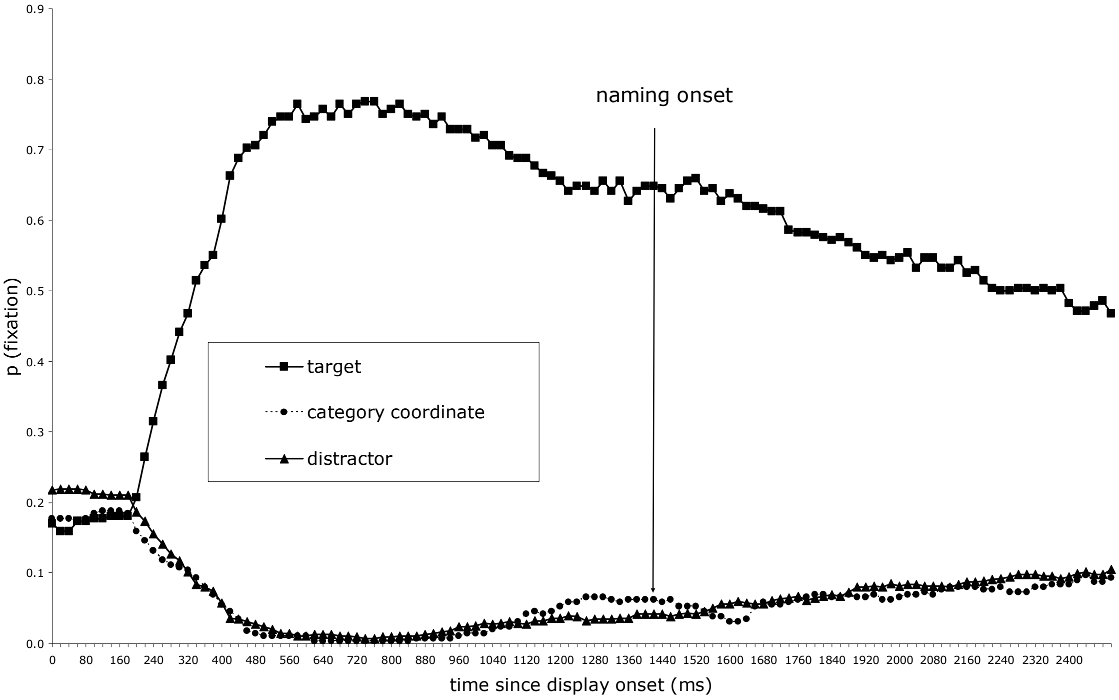


Figure 2b


p (fixation)
naming onset
target
category coordinate
distractor
time since display onset (ms)