Published in final edited form as:
J Mem Lang. 2006 May ; 54(4): 515–540.

Halting in Single Word Production: A Test of the Perceptual Loop Theory of Speech Monitoring

L. Robert Slevc and **Victor S. Ferreira**
University of California, San Diego

Abstract

The *perceptual loop theory* of speech monitoring (Levelt, 1983) claims that inner and overt speech are monitored by the comprehension system, which detects errors by comparing the comprehension of formulated utterances to originally intended utterances. To test the perceptual loop monitor, speakers named pictures and sometimes attempted to halt speech in response to auditory (Experiments 1 and 3) or visual (Experiments 2, 4, and 5) words that differed from the picture name. These *stop-signal* words were varied in terms of their semantic or phonological similarity to the intended word. The ability to halt word production was sensitive to phonological similarity and, in Experiment 5, to emotional valence, but not to semantic similarity. These results suggest that the perceptual loop detects errors by making comparisons at a level where phonological knowledge is represented. These data also imply that dialogue, back channeling, and other areas where speech production is affected by simultaneous comprehension may operate based on phonological comparisons.

Halting in Single Word Production: A Test of the Perceptual Loop Theory of Speech Monitoring

One of the more striking features of language production is its efficiency and accuracy. Levelt (1989) estimates that we produce about 150 words per minute, but make only one lexical error per 1,000 words. This is especially impressive considering the complexities of word production. When speakers produce words, they start with an idea they wish to communicate, then must retrieve both lexical and phonological information, and finally program a set of motor movements that can then be comprehended by listeners. Despite these complexities, speech production seems relatively effortless and error-free.

One reason that errors are relatively infrequent may be that speakers comprehend their own speech to inspect it for errors, thereby allowing them to inhibit and repair erroneous utterances and speak relatively fluently. It is not unusual for speakers to stop and correct themselves when they make an error, sometimes even before the error is externally apparent. This idea has been formalized as the *speech monitor*.

A number of findings in the speech-error record have been used as evidence for an inner speech monitor, such as the fact that phonological speech errors are more likely to result in real words than nonwords (the *lexical bias effect*). Baars, Motley and MacKay (1975) provided experimental evidence of this effect by using a procedure to elicit Spoonerisms (exchanges of

Please address correspondence to: L. Robert Slevc, Department of Psychology 0109, University of California, San Diego, La Jolla, CA 92093-0109, slevc@psy.ucsd.edu

the initial sounds of a pair of words), and found that exchanges that formed other words (e.g., "darn bore" from "barn door") were more likely than exchanges that formed nonwords (e.g., "dorn bef" from "born deaf"). Motley, Camden and Baars (1982) used the same procedure to show that exchanges that resulted in taboo words (e.g., making a Spoonerism out of "hit shed") were made less often than exchanges that did not result in taboo words (the *taboo-words effect*). Baars and colleagues proposed that these effects could result from use of an inner monitor or editor that is sensitive to both lexical status and social appropriateness, and thus would be more likely to detect and prevent articulation of an error resulting in a nonword or a taboo word than one that results in a real word or a more appropriate word. Further supporting this account, subjects showed elevated galvanic skin responses on trials where they avoided making taboo-word errors (relative to cases where they avoided making errors that would not result in taboo words), suggesting that the taboo errors were, in fact, generated internally even when they were not overtly produced.

The speech monitor has been proposed to be sensitive to more than just lexical status and social appropriateness – it has been claimed to be sensitive to a wide variety of errors, including conceptual errors, syntactic errors, lexical errors, phonemic errors, prosodic errors, morphemic errors, errors in appropriateness of speech, and errors relating to social context (Levelt, 1989;Postma, 2000). In fact, the inner-speech monitor has been proposed to be sensitive to nearly everything to which listeners are sensitive, leading to the relatively natural theory that the monitor uses the comprehension system to listen to both inner and outer speech (Hartsuiker & Kolk, 2001;Levelt, 1983,1989;Postma, 2000). This proposal is known as the *perceptual loop theory* of speech monitoring, and claims that a speaker's prearticulatory output (or phonetic plan) is processed by the language comprehension system, which allows the speaker to compare the comprehension of what he or she is about to say to what he or she originally intended to express. Speakers are also hypothesized to listen to their own overt speech, giving them another chance to catch errors through the same mechanism (and although it is too late to prevent errors at that point, they can still be corrected). Figure 1 depicts the perceptual loop theory, as proposed by Levelt (1983,1989).

A number of results support the claim that speakers monitor both their inner and outer speech via self-comprehension. Speakers detect similar kinds of errors in their silent speech as in their overt speech (Dell & Repka, 1992), suggesting that the same mechanism is used for both internal and external error detection. Speakers' error rates are similar when producing silent, noise-masked, or mouthed speech (Postma & Noordanus, 1996), suggesting that this monitoring does not occur at the motor level. And not only can speakers detect many of their own errors when their overt speech is noise masked, but they actually do so more quickly (on average) than when they can hear their own speech (Lackner & Tuller, 1979), suggesting that monitoring of inner speech is faster than monitoring of external speech. Speakers also show evidence of capacity restrictions on error detection and correction (Oomen & Postma, 2002;Postma, 1997), suggesting that speech monitoring is a centrally regulated (or controlled) process, which is consistent with the perceptual-loop theory (Postma, 2000). Finally, there is at least some evidence of a link between disordered speech comprehension and monitoring deficits (e.g., Marshall & Tompkins, 1982), which fits with the idea that monitoring is carried out through the comprehension system.

There is, however, some evidence that is inconsistent with the idea that monitoring relies on the perception of self-produced speech. One type of discrepant evidence comes from language-impaired populations. If monitoring relies on the comprehension system, one would expect patients with neurological impairments in comprehension to have difficulty monitoring, and patients with impairments in monitoring (such as jargon aphasics, who often are unaware of their speech errors despite suffering from severe anomia) to have corresponding problems in comprehension. However, a number of studies have found dissociations in these processes in

neurologically damaged patients (e.g., Schlenk, Huber & Wilmes, 1987;McNamara, Obler, Au, Durso & Albert, 1992;Maher, Rothi & Heilman, 1994). Hartsuiker and Kolk (2001) point out that these findings are not necessarily evidence against the perceptual loop theory. Patients with good comprehension ability but poor monitoring skills may have deficits in monitoring sub-processes that do not directly affect comprehension. The reverse dissociation is more difficult to dismiss, but it is possible that patients with intact monitoring skills and poor comprehension ability (e.g., Marshall, Rappaport & Garcia-Bunuel, 1985) might suffer from comprehension deficits primarily at levels that leave stages crucial to monitoring relatively intact.

What is perhaps a more important criticism of the perceptual loop theory (and of monitoring theories in general) is that it is not constrained enough to easily generate testable predictions (Dell & Reich, 1981;Goldrick & Rapp, 2002;Martin, Weisberg & Saffran, 1989). It is not clear exactly what elements of speech are monitored, when this monitoring occurs, and how production reacts to such monitoring. Furthermore, although the inner monitor has been proposed to be sensitive to all sorts of possible errors, it seems that speakers do not always monitor all of these dimensions, and many theories propose that the editor can be more or less sensitive depending on attentional state (e.g., Oomen & Postma, 2002;Postma, 1997) or strategic factors (Baars et al., 1975;Hartsuiker, Corley, & Martensen, 2005). Together, this makes it very difficult to generate specific predictions or to evaluate post hoc explanations, since the monitor is potentially, but not necessarily, sensitive to everything that can be comprehended.

We can begin to constrain the monitoring process by considering the architecture of the speech production system and the patterns of errors that are claimed to result from monitor operation. There is a consensus in the field of language production on the general architecture of the production system. Most theories propose multistage models where lexical processing flows from the level of lexical concepts to an intermediate level of grammatical processing involving representations called *lemmas*, then to a level of phonological processing involving representations sometimes called *lexemes*, then finally to a level of phonetic planning and articulation (e.g., Dell, 1986;Garrett, 1975;Levelt, 1989;Levelt, Roelofs, and Meyer, 1999, but see Caramazza, 1997). The existence of these different levels of lexical representation implies that a lexical error could be the result of a substitution at the lemma level or of a substitution at the lexeme level. An error that occurs during lemma selection should be a semantic error, such as substituting the word “dog” for the intended word “cat.” This type of error arises because lemmas are organized semantically, such that the activation of a particular target’s semantic representation will activate not only the target’s lemma (e.g., for “cat,”) but also lemmas that share semantic features (such as the lemma for “dog”). An error that occurs during the second stage of speech should be a lexical-phonological error, such as substituting the word “car” for the intended word “cat” (these are sometimes called *Malapropisms*). This type of error arises because lexemes are organized phonologically, so that when a particular target lexeme (e.g., “cat”) becomes active, so too will other lexemes that share phonological features (e.g., “car”). This suggests that the chance of making an error that is both semantically and phonologically similar to the intended word (a *mixed error*, such as substituting the word “rat” for the intended word “cat”) should be the sum of the probability of making a semantic error that is phonologically similar to the target and the probability of making a phonological error that is semantically similar to the target (plus some very small correction for the chance of making an error at both the semantic and phonological level). However, semantic errors are phonologically similar to the target word far more often than would be expected by chance, a finding termed the *mixed error effect* (Dell & Reich, 1981;Martin, Gagnon, Schwartz, Dell & Saffran, 1996).

One way to explain the mixed error effect is to assume that speakers' ability to detect an error in their self-produced speech (i.e., the monitor's ability to detect an error) is affected by the similarity of the error to the intended word. Thus when a semantic error is also phonologically similar to the intended utterance, or when a phonological error is also semantically similar to the intended utterance, the monitor is more likely to think that the erroneous outcome is in fact correct. The monitor is thus less likely to detect such doubly-similar errors, and so they are more common than might otherwise be expected. It is worth noting, however, that an inner monitor is only required to explain the mixed error effect (and the lexical bias effect) in models of speech production that prohibit feedback (e.g., Levelt et al., 1999). In theories that allow feedback from phonological levels to lexical levels (e.g., Dell, 1986;Dell & Reich, 1981;Rapp & Goldrick, 2000) these effects can be explained without relying on an inner monitor (see Humphreys, 2002, for evidence that the lexical bias effect is better explained by feedback than by monitoring, but see Hartsuiker et al., 2005, for evidence of strategic effects on lexical bias). Thus, a critical assumption of the monitor-based explanation of the mixed error effect (and of monitor-based explanations of error detection in general) is that the monitor is less able to detect erroneous outcomes that are more similar to the intended utterance than outcomes that are less similar. This assumption, termed here *similarity-based vulnerability*, provides a way to assess and thereby constrain monitor function by determining what kinds of similarity the monitor is, in fact, vulnerable to.

The experiments reported below aimed to assess the potential similarity-based vulnerability of the perceptual loop monitor in a tightly controlled laboratory procedure. All experiments used the same basic methodology. The task was an adaptation of what is termed a *stop-signal paradigm*, in which subjects engage in a primary task, such as a simple discrimination task, and are occasionally presented with a stop-signal (e.g., a tone) telling them to stop their response on the primary task (for a comprehensive review of the stop-signal paradigm, see Logan, 1994). The stop-signal literature is primarily concerned with identifying the conditions under which people can halt an impending response before execution. From this perspective, the first step of monitor-based editing of speech production is essentially a stop-signal task: The monitor must determine whether a difference exists between intended and monitored speech, and if so, halt the impending production before articulation.

To assess monitor function in a stop-signal paradigm, we had subjects name pictures, starting the process of message generation and word production. Sometimes a word was (externally) presented that was either the name of the picture (a *go-signal*) or that was not the name of the picture (a *stop-signal*). Subjects' task was to name the pictures quickly, but to try to stop their response if they were presented with a word that was different from the word they were in the process of producing (the picture name). Thus, just as in actual perceptual-loop function, subjects' task was to determine whether a difference existed between a to-be-produced word and a comprehended word, and if so, to halt their impending production response before articulation. The general task model for these experiments is shown in Figure 2.

If this task successfully taps into monitor functioning, then we should observe that it exhibits similarity-based vulnerability: Subjects should find it harder to halt a word-production response when the stop signals are similar to the to-be-produced words, compared to when they are dissimilar. However, we can take the task and logic further, and use it to assess how stopping performance varies as a function of the type of similarity of the stop-signal word to the picture name.

One possibility is that when monitoring comprehended speech, speakers are primarily sensitive to conceptual content. This would be so if monitoring operates by comparing comprehended conceptual representations to to-be-expressed conceptual representations, as is the case in the standard perceptual loop theory (see Figure 1; Levelt, 1983,1989). A conceptual level

comparison allows monitoring processes to compare across a 'common code' (i.e., conceptual representations, which are likely to be shared between comprehension and production processes), rather than translating across different modalities (e.g., if monitoring is performed at an acoustic and articulatory level, where representations may differ between comprehension and production). If the standard perceptual loop theory is correct, and given the notion of similarity-based vulnerability, then it should be particularly difficult to detect errors that are semantically similar to intended words. This predicts that stopping performance in the task used here should be vulnerable primarily to semantic similarity – speakers should have difficulty halting production when the comprehended word and the to-be-produced word are semantically similar.

Another possibility is that speakers are sensitive primarily to the phonological content of their own speech. This would be so if monitoring operates by comparing comprehended phonological representations to to-be-expressed phonological representations. This fits with the proposal that the perceptual loop monitor has difficulty detecting phonologically similar errors because they are in the comprehension cohort of the intended word (e.g., Marslen-Wilson & Welsh, 1978) and thus the monitor is especially vulnerable to phonological similarity (Roelofs, 2004). Given the notion of similarity-based vulnerability, this predicts that if the stop-signal is phonologically similar to the target, it should be more difficult to stop production than when it is phonologically dissimilar.

Of course, speakers might be sensitive to both the semantics and phonology of their self-produced speech, if monitoring operates by comparing comprehended to produced speech both at semantic and phonological levels of representation. If so, then it should be most difficult to stop when the signal is both semantically and phonologically similar, less difficult when the stop-signal is only semantically or only phonologically similar, and relatively easy when the stop-signal is dissimilar to the picture name.

It is worth noting that performance in this task promises insights beyond its implications for the perceptual-loop theory of speech monitoring, including with regard to the disruptive effect of *delayed auditory feedback* (e.g., Yates, 1963) and the simultaneous comprehension and production of language in dialogue (e.g., Clark, 1996). We address these issues in the General Discussion.

Experiment 1

In Experiment 1 (as in all experiments), subjects' primary task was to name pictures. As in the stop-signal literature (see Logan, 1994), 75% of the trials were either *go-trials*, (where subjects saw only the picture) or *go-signal trials* (where subjects saw the picture and heard the name of the picture), and the other 25% of the trials were *stop-signal trials*. This makes the stop-signal rare enough that subjects do not adopt a strategy of waiting for a stop- (or go-) signal before naming the picture. In Experiment 1, the go-signals and stop-signals were all presented as auditory words.

Because monitoring-based explanations rely on the monitor's vulnerability to similarity, such that errors that are more similar to the intended utterance are more likely to "slip by" the monitor than are errors that are less similar, the stop-signals were varied in terms of their semantic and phonological similarity to the target picture name. The type of comparison that the monitor makes should correspond to the type of similarity that the monitor is sensitive to, so a monitor that operates on the basis of a semantic or phonological comparison (or both) should be sensitive to semantic or phonological similarity (or both).

Method

Participants—Fifty University of California, San Diego undergraduates participated in Experiment 1 in exchange for course credit. In all experiments, participants whose mean go-trial reaction times differed by more than 2 standard deviations from the overall mean of go-trial reaction times were not included in the analysis. In Experiment 1, data from one participant exceeded this criterion and were excluded, although the results do not change appreciably when this participant is included in the analysis. One additional participant, who misunderstood the instructions and performed the task incorrectly, was also excluded. All participants reported learning English as their native language.

Materials—The materials were adapted from Experiment 3 of Damian and Martin (1999). These were eighteen line drawings of common objects taken from Snodgrass and Vanderwart's (1980) set. Each picture had a semantically similar word that also shared a minimum of the first two phonemes with the picture name (e.g., *lantern* for a picture of a lamp). Three other words were paired with each picture: a semantically similar word that was matched to the semantically-and-phonologically similar word in terms of semantic overlap to the target picture name but was phonologically dissimilar (e.g., *candle*), a phonologically similar word that was matched to the semantically-and-phonologically similar word in terms of phonological overlap to the target picture name but was semantically dissimilar (e.g., *landing*), and a word that had no obvious similarity to either the picture name or to any of the other distracter words (e.g., *package*). Phonological similarity required an overlap of at least the two initial phonemes, and semantic similarity required a match between the rated semantic similarity of the semantically related word to that of the semantically and phonologically similar word (for further details see Damian and Martin, 1999). The stimuli are listed in Appendix A.

Design and analysis—The experimental design included two factors, each with two levels: phonological similarity (similar or dissimilar) and semantic similarity (similar or dissimilar). Both factors were varied factorially within subjects. Items were presented to each participant in random order. Stopping accuracy and response latencies (when the participant provided a response) were measured on each trial. In this and in all following experiments, trials were considered successfully stopped when a maximum of two phonemes were produced before halting. This criterion was chosen because errors halted after only two phonemes of external speech are very likely to have been detected via the internal loop (at least according to the perceptual loop theory) as the time it would take to perceive enough of an erroneous speech signal to detect an error through the outer loop would exceed the time it takes to produce two phonemes. The pattern of results is similar when other criteria are adopted (e.g., considering only completely halted trials as successfully halted, or when counting every non-completed word as successfully halted), except for one comparison, discussed in the results of Experiment 5. Stopping performance was analyzed with 2 × 2 repeated-measures analyses of variance (ANOVAs), using phonological and semantic similarity as factors and subjects ($F1$) and items ($F2$) as random variables.

Stop-signal reaction times (SSRTs) were also calculated (Logan & Cowan, 1984). An SSRT estimates the reaction time of the internal response to the stop-signal cue, so is essentially an estimate of the average time it takes speakers to perceive and evaluate the stop signal and to perform the "action" of stopping. Under the assumption that the stopping process and the primary task process (picture naming) are independent, SSRTs can be calculated by using the distribution of RTs on the go trials as an approximation of the covert distribution of RTs to the stop-signal (i.e., the distribution of SSRTs). Under a "horse race" model of the stopping process (where the primary task process and the stopping process race independently to completion; Logan & Cowan, 1984), the finish of the stopping process splits the SSRT distribution such that fast stop-responses to the stop-signal are halted and slow responses are not, thus the average

finish time of the stop-signal process can be estimated as the value of the go-trial RT distribution corresponding to the proportion of successfully halted stop-signal trials. Patterns of SSRTs, then, will be similar to the patterns of stopping performance, however they provide additional timing data that may be relevant to chronometric models of self-monitoring and production (e.g., Hartsuiker & Kolk, 2001).

Primary task response latencies (i.e., picture naming latencies) were analyzed only for trials that were not halted (i.e., for trials on which a response latency can actually be observed) and for trials where subjects accurately produced the picture name. Also, latencies that were above or below three standard deviations from each subject's mean go-trial picture-naming latency were excluded. These picture-naming latencies were analyzed in similar 2 × 2 repeated-measures ANOVAs with phonological and semantic similarity as factors.

For all analyses, variability is reported with repeated measures 95% confidence-interval half-widths (CIs) based on single degree-of-freedom comparisons, for subjects and for items (Loftus & Masson, 1994;Masson & Loftus, 2003). In all experiments, effects reported as significant are at or below the .05 alpha level unless indicated otherwise.

Apparatus—The experiment was administered on Apple Macintosh computers running PsyScope 1.2.5 (Cohen, MacWhinney, Flatt, & Provost, 1993). The pictures were approximately 3 in. by 3 in. black line drawings shown on a white background. The auditory stimuli were recorded by the first author, digitized at a sampling frequency of 44.1 kHz, and presented through a speaker next to the screen. Vocal responses were recorded using a head-worn microphone connected to a PsyScope button box (which measures voice onset latencies) and a standard cassette recorder. Responses were recorded to tape and used to verify stopping performance. The microphone sensitivity was calibrated separately for each subject.

Procedure—At the beginning of each experimental session, subjects were familiarized with the set of pictures by viewing each picture and its name on the screen for 2,000 ms. A practice set of trials was administered where subjects saw the eighteen pictures presented in a random order, and were required to provide the correct picture name for each. Any incorrect responses were corrected by the experimenter. Subjects then saw interactive instructions on the computer screen and heard the experimenter summarize the task.

Subjects were told that their main task was to name pictures as soon as they could, but that they would occasionally need to try to stop their naming response and say nothing. On 37.5% of the trials (the go trials), the subjects saw a picture and did not hear any word, in which case they were to simply name the picture. On another 37.5% of the trials (the go-signal trials), the name of the picture was presented shortly after the picture appeared, in which case subjects again were to name the picture. On the remaining 25% of trials (the stop-signal trials), subjects saw the picture then heard a word that was not the picture name, and were to attempt to stop their naming response and say nothing. The experimenter stressed that it was important to name the pictures quickly, and not wait to determine if a word was presented before starting to name the picture.

Subjects were given 144 practice trials (54 go trials, 54 go-signal trials, and 36 stop-signal trials), with a break halfway through. In all experiments, stop-signals in the practice trials were unrelated to the picture names. On each trial, subjects saw the trial number presented for 500 ms in the center of the screen, which was replaced with a fixation cross for 500 ms, followed by the picture. On trials with auditory stop- or go-signals, the signal was presented after a short delay (the stimulus onset asynchrony, or SOA). In every case, the picture remained on the screen until the voice key triggered or 2000 ms passed, when the picture disappeared and the next trial started.

During practice, the SOA was initially set at 400 ms and was varied according to stop-task performance. Each time a subject was able to stop on a stop-signal trial, the SOA was increased by 10 ms, making the task slightly more difficult, and each time a subject failed to stop on a stop-signal trial, the SOA was decreased by 10 ms, making the task slightly easier. This method of calibrating the stop-signal delay, similar to that proposed by Osman, Kornblum, and Meyer (1986), provides a greater amount of data than a static SOA since it can adjust for individual differences in stopping ability. However, the SOA was not allowed to drop below 280 ms in order to reduce the likelihood that the words could speed or slow the picture naming responses (as in picture-word interference paradigms, e.g., Schriefers, Meyer, & Levelt, 1990; this issue is discussed further below).

Subjects were then presented with 288 experimental trials: 108 go trials (each picture 6 times with no signal), 108 go-signal trials (each picture 6 times with the picture name as a go-signal), and 72 stop-signal trials (each picture 4 times, once with each type of stop-signal). These were the same as the practice trials, except that the SOA was held constant at the level reached at the end of the practice, and the stop-signal words were varied according to the factors of semantic and phonological similarity. The trials were presented in random order, and subjects were given two breaks, equally spaced throughout the experimental session. Including instructions and practice, the experimental session lasted approximately 35 minutes.

Results

Figure 3 shows the mean stopping accuracy as a function of semantic and phonological similarity of the stop-signal to the picture name. Subjects successfully stopped their naming response on an average of 38.7% of stop-signal trials in the phonologically dissimilar conditions, but only on an average of 20.8% of stop-signal trials in the phonologically similar conditions. There was no difference between stopping accuracy in the semantically dissimilar (30.2% of stop-signal trials) and similar (29.3% of stop-signal trials) conditions, and the effect of phonological similarity was the same when stop signals were semantically similar (a 17% difference) compared to when they were semantically dissimilar (a 19% difference). These observations were supported by statistical analyses. The main effect of phonological similarity was significant ($F1(1,47) = 81.2$, CI = ±4.0%; $F2(1,17) = 153.3$, CI = ±3.1%), the main effect of semantic similarity was not significant ($F1(1,47) < 1$, CI = ±2.4%; $F2(1,17) < 1$, CI = ±3.1%), and the interaction between phonological and semantic similarity was also not significant ($F1(1,47) < 1$, CI = ±3.7%; $F2(1,17) < 1$, CI = ±3.8%).

Expressed in terms of Stop Signal Reaction Times (SSRTs), this stopping performance corresponds to a slower estimated SSRT to phonologically similar stop-signals than to phonologically dissimilar stop-signals (356 ms vs. 284 ms; $F(1,47) = 66.0$, CI = ±18.0 ms), but similar estimated SSRTs to semantically similar and dissimilar stop signals (320.0 ms vs. 320.3 ms; $F(1,47) < 1$, CI = ±16.0 ms).[1]

Table 1 lists the mean picture-naming reaction times (RTs) for each type of trial in Experiments 1 and 2. The mean RTs in the stop-signal trials are slightly faster than in the go and go-signal trials. This is probably because the stop-signal trial RTs are only taken from the trials on which subjects failed to inhibit their response, which will generally be the faster part of the RT distribution because the trials with slower RTs would be more likely to have been successfully stopped (as is assumed for the computation of SSRTs). In Experiment 1, mean naming latencies (when subjects were unable to stop) in the phonologically similar conditions were significantly faster than in the phonologically dissimilar conditions (534 ms vs. 554 ms; $F1(1,47) = 8.23$,

[1]Note that it was not possible to accurately estimate SSRTs by items in Experiments 1-3 because the SOA varied across subjects. Thus SSRT analyses were only conducted using subjects as random variables in these experiments.

CI = ±13.7 ms; $F2(1,17) = 12.7$, CI = ±9.3 ms). RTs in the semantically similar and dissimilar conditions were not significantly different (540 ms vs. 548 ms; $F1(1,47) = 2.11$, CI = ±11.8 ms; $F2(1,17) = 1.99$, CI = ±10.3 ms) and there was no significant interaction between phonological and semantic relatedness on picture naming latencies ($F1(1,47) = 1.28$, CI = ±11.5 ms; $F2(1,17) = 1.81$, CI = ±10.8 ms).

Discussion

Experiment 1 found that subjects had a harder time halting production of a picture name in response to a stop-signal that was phonologically similar to the picture name than in response to a stop-signal that was phonologically dissimilar. However, subjects had no more difficulty halting production in response to a stop-signal that was semantically similar to the picture name than to a stop-signal that was semantically dissimilar. Under the assumption that the present task reveals monitor function, this pattern of stopping performance provides support for an inner monitor that makes phonologically based comparisons, but that is insensitive to semantic information.

Subjects found the task difficult, as shown by the relatively poor stopping performance even in the phonologically dissimilar conditions. Because of the SOA-calibration procedure, performance on the phonologically and semantically dissimilar stop-signal trials was expected to be around 50%. However, subjects were only able to stop on an average of 39% of those trials. This may be due to the floor imposed on the SOA, making the calibration imperfect, or perhaps the SOA calibration did not go on long enough to stabilize. Note, however, that this worse-than-expected stopping performance does not alter the main conclusion that subjects found it more difficult to halt in response to phonologically similar stop-signals than to phonologically dissimilar stop-signals. This also suggests that subjects were not waiting for the stop (or go) signals before they began their responses.

There are two more important concerns with these results. One is the possibility that the pattern of stopping accuracy resulted from a speed-accuracy tradeoff, because the conditions with the lowest stopping accuracy were also the trials with the fastest naming times. We address this concern below, in light of the results of Experiments 2–4. Another concern is the possibility that the ineffectiveness of the phonologically similar stop-signals (relative to the phonologically dissimilar stop-signals) was due to a strategic effect. In particular, it may be that subjects employed a strategy of simply listening for any discrepancy between the picture name and the signal, and made the decision to halt speech based on the first detected discrepancy. Although the onset of the words in each condition occurred with the same SOA, the first point of discrepant information differs across conditions. Since all words in the phonologically similar conditions were similar in onset to the picture name (and therefore to the go-signal), it is not until part of the way through the word that the discrepant information becomes available. For example, if the participant is trying to name the picture of a lamp, as soon as they begin to hear the /p/ of /pækɪdʒ/ (package) or /k/ of /kændl/ (candle) they can immediately determine that the word they are hearing is not /læmp/ (lamp). However, when they hear /lændŋ/ (landing) or /læntərn/ (lantern), it is not until the third phoneme (ignoring coarticulation effects) that they are able to determine that the word they are hearing is not “lamp.” Thus, one could argue that the time between the picture onset and the stop-signal is effectively *longer* (i.e., harder) in the phonologically similar conditions than in any of the phonologically dissimilar conditions. It is important to note, however, that this concern is still relevant to speech monitoring; because picture stimuli do not encode their names, speakers must have compared the incoming stop signals to internally generated representations of the picture names (see Wheeldon & Levelt, 1995, and Wheeldon & Morgan, 2002, for investigations of monitoring that rely on this logic).

There are two ways to interpret this latter concern. One is as a task artifact: The auditory nature of the stop-signals in Experiment 1 caused those stop-signals to unfold in time, leading to the later arrival of discrepant information only in the phonologically similar conditions. By this interpretation, any specific sensitivity to similarity at the beginnings of words is due solely to the necessarily later arrival of discrepant information in this task, and it does not reflect anything about natural production. Experiments 2–4 were designed to address this possibility. The second interpretation is that the unfolding nature of the stop-signals actually corresponds to the unfolding nature of the inner-speech signal. That is, just as an externally presented stop-signal unfolds in time, leading to the later arrival of phonologically discrepant information for beginning-similar words, so too might the inner-speech stream unfold in time, leading to the later arrival of phonologically discrepant information, but only for beginning-similar words. In turn, this would predict that in natural speech errors, lexical errors that are phonologically similar to intended words (Malapropisms) should be sensitive to the position of similarity between actual and intended words, such that outcomes that are more similar at the beginnings of words should be more prevalent than outcomes that are similar at the ends. In fact, an analysis of naturally occurring word-substitution errors (Dell & Reich, 1981) and an analysis of picture naming errors (Martin et al., 1996) provide evidence consistent with this possibility, as both showed that erroneous outcomes and intended outcomes were most likely to be phonologically similar at the beginning of words and that phonological similarity tended to decrease deeper into the words (see Figure 2 in Dell & Reich, 1981, and Table 2 in Martin et al., 1996). We discuss this latter interpretation further below in the Discussion of Experiment 2.

Experiment 2

An important concern with Experiment 1 is that stopping performance may have had little to do with editor performance, but instead was an artifact of the later availability of discrepant information in the phonologically similar conditions. If this effect drove the differences in stopping performance for phonologically similar and dissimilar stop-signals in Experiment 1, then there should be no differences in stopping performance on a version of the task where all information, discrepant and otherwise, is available at the same time in each condition.

Experiment 2 aimed to test this possibility by using visual stop-signals. If the effect of phonological similarity found in Experiment 1 was based on a strategy of waiting for the first point of discrepant information, then the differences between phonologically similar and dissimilar stop-signals should disappear because, with visual signals, the entire word is presented at the same time. That is, if performance is completely strategically driven, speakers could process the ends of the visually presented stop-signals immediately upon their presentation, and therefore detect a discrepancy in all stop-signal conditions equally rapidly. Also, the concern that the pattern of results in Experiment 1 was due to a speed-accuracy tradeoff would be lessened if the picture naming times in Experiment 2 were found to be unrelated to stopping accuracy.

Method

Participants—Fifty-one University of California, San Diego undergraduates participated in Experiment 2 in exchange for course credit. One participant was excluded from the analysis for exceeding the go-trial RT threshold, although the results do not change appreciably if this participant is included in the analysis. Two additional participants were excluded from analysis because post-experiment discussion revealed that they misunderstood the instructions and performed the task incorrectly. All participants reported English as their native language.

Materials, Design, and Procedure—The experimental design and materials were identical to those used in Experiment 1 (see Appendix A). The procedure was also identical to that of Experiment 1, with the exception of how the stop and go signals were presented. On

each stop- and go-signal trial, the signal word was presented visually after the SOA, in uppercase Helvetica 18-point bold font in the center of each picture. The word was presented for 200 ms, then replaced with a stimulus mask ("XXXXXXX"). The picture and mask remained on the screen until the voice key was triggered, or for 2,000 ms.

Results

Figure 4 shows the mean stopping accuracy as a function of phonological and semantic similarity of the stop-signal to the picture name. The general pattern of results replicated those of Experiment 1, though with smaller differences. Subjects successfully stopped more often in the phonologically dissimilar conditions (37.2% of stop-signal trials) than in the phonologically similar conditions (32.2% of stop-signal trials), but were no more successful stopping in the semantically dissimilar conditions than in the semantically similar conditions (34.8% and 34.5% of stop-signal trials, respectively). These observations were confirmed by statistical analyses, which showed a significant main effect of phonological similarity ($F1(1,47) = 7.60$, CI = ±3.7%; $F2(1,17) = 9.54$, CI = ±2.7%) but no effect of semantic similarity ($F1(1,47) < 1$, CI = ±2.9%; $F2(1,17) < 1$, CI = ±3.1%) and no interaction between phonological and semantic similarity ($F1(1,47) = 2.28$, CI = ±3.7%; $F2(1,17) = 2.20$, CI = ±4.4%). Expressed in terms of estimated SSRTs, speakers were slower to stop in response to phonologically similar than to phonologically dissimilar stop-signals (303 ms vs. 286 ms; $F(1,47) = 10.6$, CI = ±10.6 ms), but no different in response to semantically similar and dissimilar stop-signals (294 ms vs. 295 ms; $F(1,47) < 1$, CI = ±10.0 ms).

The picture-naming latencies for Experiment 2 are reported in Table 1. On stop-signal trials, there were no differences between RTs in the phonologically similar and dissimilar conditions (554 ms vs. 550 ms; $F1(1,47) < 1$, CI = ±7.3 ms; $F2(1,17) < 1$, CI = ±6.7 ms), nor were there differences between RTs in the semantically similar and dissimilar conditions (553 ms vs. 551 ms; $F1(1,47) < 1$, CI = ±6.7 ms; $F2(1,17) < 1$, CI = ±6.3 ms), and these two factors did not interact ($F1(1,47) = 2.08$, CI = ±9.2 ms; $F2(1,17) = 2.08$, CI = ±12.1 ms).

Discussion

Experiment 2 provides further evidence that, as diagnosed by the current task, the inner monitor is sensitive to phonological similarity and not semantic similarity, as even with visually presented stop-signals, subjects found it harder to stop when stop-signals were phonologically similar to the intended words but not when they were semantically similar to the intended words. That this pattern emerged with visual stop-signals also suggests that the effect is not fully determined by a strategy of waiting for discrepant information before halting a word-production response.

Note that the phonological similarity effect was considerably larger in Experiment 1 than in Experiment 2 (18% vs. 5%). This observation is supported by a direct statistical comparison between Experiments 1 and 2 (treating Experiment as a between-subjects factor), which revealed an overall significant main effect of phonological similarity ($F1(1,94) = 72.7$, CI = ±3.8%; $F2(1,17) = 156$, CI = ±3.9%) and, more importantly, a significant interaction between phonological similarity and Experiment ($F1(1,94) = 23.2$, CI = ±3.8%; $F2(1,17) = 28.3$, CI = ±5.2%), but no other significant effects. This shows that the effect of phonological similarity was significantly greater with stop-signals in the auditory modality, when the availability of discrepant information was relatively delayed (as in Experiment 1), compared to stop-signals in the visual modality, when similar and discrepant information were available simultaneously (as in Experiment 2).

It is still possible, however, that the discrepant information arrives later in the phonologically similar conditions – even with a visually presented stop-signal – if words are read (i.e., visually

processed) in English from left to right rather than as a single unit. In fact, theories of reading (e.g., Rayner & Pollatsek, 1989;Coltheart, Rastle, Perry, Langdon & Ziegler, 2001) suggest that low frequency and long words are processed serially, from left to right, while high frequency and short words are processed in parallel over the length of the word, and thus as single units. This predicts that if the left-to-right nature of word recognition drove the results of Experiment 2, phonological similarity effects should have been greater for longer and lower frequency words than for shorter and higher frequency words. However, regression analyses showed that the (logarithm transformed) frequency of the stop-signal words, calculated using the CELEX lexical database (Baayen, Piepenbrock, & van Rijn, 1993), did not predict stopping performance overall ($r = .15$, $F(1,71) = 1.61$), nor just in the phonologically similar conditions ($r = .11$, $F(1,35) < 1$). The length of the stop-signal words also did not predict stopping performance overall ($r = .04$, $F(1,71) < 1$), nor did it predict performance in the phonologically similar conditions ($r = .06$, $F(1,35) < 1$). Thus we have no evidence, assuming more parallel comprehension of short or high frequency words, that difficulty in stopping is related to phonologically decoding the stop-signals from left-to-right. This suggests that the difference in stopping performance between the phonologically similar and dissimilar conditions may be due to the overall phonological similarity of the stop-signal to the picture name, rather than just to any phonological similarity at the beginning of the stop-signals.

Speech Error Analysis

If it is true that the effect of visually presented phonologically similar stop-signals in Experiment 2 demonstrates an effect of overall phonological similarity, then we should see that naturally occurring Malapropisms do not tend to be more similar to intended words at the beginnings of words than elsewhere. As noted above, the findings of Dell and Reich (1981) and of Martin et al. (1996) suggest the opposite: They found that the phonological similarity of intended and erroneous outcomes (in naturally occurring word-substitution errors and in elicited picture-naming errors, respectively) seemed to be strongest on the initial phoneme, which implies that phonological order may, in fact, be relevant in error detection. However, measuring phonological similarity as the percentage of matching phonemes in each phoneme position calculated from the beginning of the word is problematic because two given words are likely to become more structurally misaligned as one proceeds deeper into a word, which makes similarity more likely to be missed later in the words (e.g., if the word / trɪk/ (trick) was erroneously produced as / tɪk/ (tick), this measure would inaccurately suggest that this error was phonologically similar only on the first phoneme, because that is the only position of overlap when counted from the beginning of the words). Martin et al. (1996) addressed this concern by also looking only at the first stressed syllable, and found the same pattern with phonological overlap most likely on the initial consonant, less likely on the stressed vowel, and even less likely on the final consonant. Still, it is unclear if this effect is specific to the first stressed syllable, or if the likelihood of phonological overlap drops off over the entire length of the word. A simple way to address this issue is to explore the overlap of both the first phoneme and the last phoneme of the actual-intended pairs. If errors are, in fact, more likely to be beginning-similar, then the first phoneme of actual-intended word pairs should be phonologically similar more often than predicted by chance, but the last phoneme of actual-intended word pairs should only show chance levels of similarity.

To investigate this, we conducted an analysis of 274 Malapropisms (specifically, word substitution errors where the target-error pairs were both grammatically and formally related) from the UCLA Speech Error Corpus (UCLA Speech Error Corpus [Data file]; Fromkin, 1971; this corpus is currently available online: http://www.mpi.nl/corpus/sedb/index.html), all of which involved content words and were cases where the intended word was clearly specified. We calculated the percentage of phonemes that overlapped between actual and intended words both on the first phoneme and on the last phoneme (essentially by right-aligning the actual-

intended pairs and calculating the percentage of last phonemes that matched). We then calculated chance estimates of phonological overlap on both the first and last phoneme by randomly re-pairing the actual and intended words 100 times and calculating the percentage of phonological overlap between the actual and intended words in each re-pairing. The points on the left side of Figure 5 show a similar result as shown in Dell and Reich (1981): the percentage of first phonemes that overlap between the intended and actual words was quite high (61.3%) and considerably higher than any of the 100 chance estimates (mean = 6.4%). Interestingly, the points on the right side of Figure 5 shows that the pattern is essentially the same for the last phoneme of these errors, which show considerably more phonological overlap (65.3%) than any of the 100 chance estimates (mean = 10.3%). Thus these results suggest that errors that are phonologically similar to the beginnings or to the ends of intended words (or both) are considerably more common than would be expected by chance, and that there is not a bias for Malapropisms to be beginning-similar. This further suggests that the drop-off in phonological overlap observed in Dell and Reich (1981) and Martin, et al. (1996) was not caused by differential likelihood of phonological overlap across the length of words, but was perhaps due to structural misalignment or was specific to the first stressed syllable.

In sum, both the results of Experiment 2 with visual stop-signals and of an analysis of naturally occurring Malapropisms suggests that any vulnerability of the speech monitor to phonological similarity should arise both for stop-signals that are similar at the beginnings of words, and for stop-signals that are similar at the ends of words (i.e., rhyming words). Specifically, these observations suggest that subjects doing the same task as was used in the first two experiments, but with rhyming stop-signals, should have comparable difficulty with the rhyme related stop-signals as with the onset related stop-signals in the first two experiments. Experiment 3 was designed to examine this possibility, and to replicate the ineffectiveness of the semantically similar versus dissimilar stop-signals with a new set of items.

Experiment 3

Experiments 1 and 2 showed that phonological similarity but not semantic similarity affected speakers' ability to halt speech. However, the phonologically related signals in these experiments were all similar in onset to the picture name. Another way to determine whether the effect is specific to onset-similar words is to see if this pattern generalizes to signals that are phonologically similar in a way that is not based on onset. Experiments 3 and 4 used a different set of materials, in which the phonologically similar condition consisted of stop-signals that rhymed with the picture name (and thus the intended word) to address this issue. Also, Experiments 3 and 4 did not test stop signals that were similar both semantically and phonologically to picture names, due to their limited availability; instead, we manipulated semantic and phonological similarity separately.

Method

Participants—Fifty University of California, San Diego undergraduates participated in Experiment 3 in exchange for course credit. Two participants were excluded from the analyses because of equipment malfunction. All participants reported English as their native language.

Materials—Twenty-four line drawings of common objects were chosen from Snodgrass and Vanderwart's (1980) set. For each picture, a semantically similar word and a phonologically similar word (one that rhymed with the picture name) were chosen. Unrelated conditions were created by counterbalanced re-assignment of the similar words, to control for possible idiosyncratic effects of particular stop-signals. Thus, each stop-signal word appeared twice – once with the picture to which it was either semantically or phonologically similar and once with a picture to which it bore no obvious semantic or phonological similarity. The stimuli for Experiments 3 and 4 are listed in Appendix B.

Design, Analysis, and Procedure—The procedure was identical to that in Experiment 1 (i.e., auditorily presented stop-signals), except there was a total of 192 practice trials and 384 experimental trials. Stopping performance and picture naming latencies were analyzed with a 2 × 2 repeated-measures ANOVA design, using type-of-relationship (phonological or semantic) and similarity (similar or dissimilar) as factors. Additionally, planned comparisons were carried out to contrast the phonologically similar condition with the phonologically dissimilar condition, and to contrast the semantically similar condition with the semantically dissimilar condition.

Results

Stopping accuracies as a function of the type of stop-signal are shown in Figure 6. There was a small difference in stopping accuracy between the phonologically similar stop-signals and their controls (34.8% vs. 38.1% of stop-signal trials) and a smaller difference in stopping accuracy between the semantically similar stop-signals and their controls (37.6% vs. 40.0% of stop-signal trials), though the difference in the size of these differences is negligible. Statistical analyses showed a significant main effect of similarity ($F1(1,47) = 4.23$, CI = ±2.8%; $F2(1,23) = 5.42$, CI = ±2.3%) and an effect of type-of-relationship that was significant by subjects ($F1(1,47) = 4.27$, CI = ±2.2%), but not by items ($F2(1,23) = 2.23$, CI = ±3.0%). Although the phonological difference was numerically larger than the semantic difference (consistent with the results of Experiments 1 and 2), the interaction between these factors was not significant ($F1(1,47) < 1$, CI = ±3.4%; $F2(1,23) < 1$, CI = ±3.8%). Finally, planned comparisons revealed that the difference between the phonologically similar and phonologically dissimilar conditions was marginally significant ($F1(1,47) = 3.65$, $p < .07$; $F2(1,11) = 3.12$, $p < .1$), whereas the difference between stopping accuracy in the semantically similar and semantically dissimilar conditions was not ($F1(1,47) = 1.97$; $F2(1,11) = 1.83$).

Expressed in terms of estimated SSRTs, this stopping performance corresponds to a slower SSRT to phonologically similar stop-signals than to their controls (254 ms vs. 227 ms), and a slightly slower SSRT to semantically similar stop-signals than to their controls (240 ms vs. 237 ms). Statistical analysis of the SSRT data showed a significant effect of type-of-relationship ($F(1,47) = 4.63$, CI = ±7.7 ms), but no main effect of similarity ($F(1,47) = 2.01$, CI = ±12.3 ms) and no interaction between these factors ($F(1,47) = 1.23$, CI = ±12.4 ms). As above, planned comparisons showed a significant difference between the phonologically similar condition and its dissimilar control ($F(1,47) = 4.54$), but no difference between the semantically similar condition and its dissimilar control ($F(1,47) < 1$).

The picture-naming latencies for Experiment 3 are reported in Table 2 (note that one subject was not included in this RT analysis because of missing values that occurred when all responses in a given condition were successfully halted). There was no significant main effect of similarity ($F1(1,46) < 1$, CI = ±7.2 ms; $F2(1,23) < 1$, CI = ±8.7 ms), but responses were slightly faster to the related stop-signals than to their controls (515.9 ms vs. 523.7 ms respectively), a significant main effect of type-of-relationship by subjects ($F1(1,46) = 4.30$, CI = ±7.2 ms), but not by items ($F2(1,23) = 1.89$, CI = ±9.8 ms), and there was no interaction between the two factors ($F1(1,46) < 1$, CI = ±11.6 ms; $F2(1,23) = 2.23$, CI = ±11.8 ms).

Discussion

Although the pattern of results in Experiment 3 is qualitatively similar to that in Experiments 1 and 2, the difference between stopping performance in the rhyme-similar and dissimilar conditions was only marginally significant. Most relevant is the comparison to Experiment 1, since both Experiment 1 and Experiment 3 used auditory stop-signals. One way to explain why the effect of phonological similarity was less robust in Experiment 3 is to consider two factors that may have influenced stopping performance in Experiment 1. First, the ability to detect that

a comprehended word is different from a to-be-produced word may be vulnerable to whole-word phonological similarity, as suggested by the results of Experiment 2 and the analysis of naturally occurring Malapropisms. Such an effect should manifest in all experiments. Second, the ability to detect a difference may be additionally affected by the later arrival of phonologically dissimilar information with auditorily presented onset-related signals, which would be observed only in Experiment 1. This second factor is likely to have had the opposite effect in Experiments 1 and 3: because the onset of all stop-signals were phonologically dissimilar in Experiment 3, speakers may often have already initiated the signal to halt in the phonologically similar conditions before the phonologically related material had been processed. The presence of a whole-word effect in Experiment 2 led to a 5% effect of phonological similarity that was significant, and a 3.3% effect in Experiment 3 that was marginally significant, compared to the 18% (significant) effect observed in Experiment 1. In short, all three experiments thus far may have revealed an effect of phonological similarity on speakers' ability to detect that a comprehended word is different from a to-be-produced word, but this effect in Experiments 1 and 3 was affected by the temporal properties of auditory stop-signals. Specifically, this effect in Experiment 1 was exaggerated by the (possibly strategically relevant) influence of the later arrival of phonologically dissimilar information with onset-similar auditorily presented stop-signals, whereas in Experiment 3 this effect was reduced by the influence of the later arrival of phonologically similar information with rhyme-similar auditorily presented stop-signals.

If the rhyming stop-signals used in Experiment 3 do show the whole-word effect of phonological similarity, like the visually presented onset-related signals in Experiment 2, then rhyming signals that are presented visually should still cause that same whole-word effect of phonological similarity. Experiment 4 tested this prediction. Furthermore, note that the difference between phonological similarity conditions in Experiment 2 was larger than in Experiment 3, hinting that with respect to the whole-word effect, visual stop-signals may be more effective than auditory stop-signals (below, we speculate about why this may be). Additionally, performance with the semantically related stop-signals in Experiment 3 showed a numerical difference consistent with the possibility that semantic similarity causes difficulty in halting performance (which potentially compromised observing an interaction between similarity type and relatedness). If that effect of semantic similarity was real, then it should be observed with the visually presented stop-signals presented in Experiment 4.

Experiment 4

Experiment 3 showed only a marginally significant difference between stopping performance in the rhyme-similar condition versus its control condition. If this marginally significant difference reflected a whole-word effect of phonological similarity, then that effect should be observed with visually presented rhyming stop-signals, as was observed for visually presented onset-similar signals in Experiment 2. But if the marginal difference reflected a non-effect of phonological similarity, implying that the phonological similarity effect arises only with onset-related materials, then visually presented rhyming words should show no difference in Experiment 4.

Method

Participants—Forty-eight University of California, San Diego undergraduates participated in Experiment 4 for course credit. All participants reported English as their native language.

Materials, Design, and Procedure—The experimental design and materials were identical to those used in Experiment 3 (see Appendix B). The procedure was also identical to that of Experiment 3, with two exceptions. First, on each stop- and go-signal trial, the stop- and go-signals were presented visually in uppercase Helvetica 18-point bold font in the center

of each picture for 200 ms, and then replaced with a stimulus mask ("XXXXXXX"). The picture and the mask remained on the screen until the voice key triggered, or for 2,000 ms. Second, the practice procedure was shortened by eliminating the SOA calibration phase. The SOA was held constant at 300 ms throughout the practice and the experimental trials, which was chosen because, for a majority of the subjects in the previous three experiments (124 of 144 subjects, or 86%), the calibration phase resulted in an SOA of approximately 300 ms.

Results

The results from Experiment 4 are shown in Figure 7. Subjects were less successful halting speech when the stop-signals were phonologically similar (in rhyme) to the picture names (36.3% of stop-signal trials) than when those same stop-signals were phonologically dissimilar to the picture names (42.5% of stop-signal trials). Subjects had no more trouble halting speech when the stop-signals were semantically similar than when those same stop-signals were semantically dissimilar to the picture names (41.3% and 41.0% of stop-signal trials respectively). Statistical analyses confirm these observations. A significant main effect of similarity was observed by subjects ($F1(1,47) = 8.03$, CI = ±2.2%) and was marginally significant by items ($F2(1,23) = 4.11$, CI = ±3.3%, $p < .06$) and a marginally significant main effect of type-of-relationship was observed by subjects ($F1(1,47) = 3.37$, CI = ±1.8%, $p < .08$) but not by items ($F2(1,23) = 2.92$, CI = ±1.9%). Most importantly, a significant interaction between similarity and type-of-relationship was observed ($F1(1,47) = 6.66$, CI = ±3.2%; $F2(1,23) = 6.25$, CI = ±3.8%). This reflected the fact that, as shown by planned comparisons, the effect of similarity within the phonological condition was significant ($F1(1,47) = 14.5$; $F2(1,11) = 13.0$), whereas the effect of similarity within the semantic condition was not ($F1(1,47) < 1$; $F2(1,11) < 1$).

Expressed in terms of estimated SSRTs, this stopping performance corresponds to a slower SSRT to phonologically similar stop-signals than to their controls (305 ms vs. 292 ms), and similar SSRTs to semantically similar stop-signals and to their controls (285 ms vs. 288 ms). Statistical analysis of the SSRT data showed a main effect of similarity, significant by subjects only ($F1(1,46) = 7.38$, CI = ±10.6 ms; $F2(1.23) = 2.41$, CI = ±10.2 ms), a main effect of type-of-relationship, significant by items only ($F1(1,46) < 1$, CI = ±11.4 ms; $F2(1,23) = 4.39$, CI = ±14.3 ms), and a significant interaction between these factors ($F1(1,46) = 3.03$, CI = ±12.9 ms; $F2(1,23) = 5.25$, CI = ±12.1 ms). As above, planned comparisons showed that this interaction reflects the fact that the difference between the phonologically similar condition and its dissimilar control was significant (although only marginally so by subjects; $F1(1,46) = 3.71$, $p < .07$; $F2(1,23) = 11.4$), but the difference between the semantically similar condition and its dissimilar control was not ($F1(1,46) < 1$, $F2(1,23) < 1$). Note that one subject was not included in the SSRT analysis because of missing values that occurred when all responses in a given condition were successfully stopped.

The picture naming latencies in Experiment 4 are reported in Table 2. Two subjects were not included in the RT analysis: one because of missing values that occurred when all responses in a given condition were successfully stopped, and one because of equipment malfunction. Latencies were slightly shorter in the semantic conditions (similar and control) than in the phonological conditions (531 ms vs. 543 ms) as indicated by a significant main effect of similarity ($F1(1,45) = 5.63$, CI = ±9.8 ms; $F2(1,23) = 5.67$, CI = ±7.7 ms) but there was no significant main effect of type-of-relationship ($F1(1,45) < 1$, CI = ±6.4 ms; $F2(1,23) < 1$, CI = ±6.6 ms) and no significant interaction between these factors ($F1(1,45) < 1$, CI = ±9.7 ms; $F2(1,23) < 1$, CI = ±9.6 ms).

Discussion

Experiment 4 showed that subjects have more difficulty halting speech in response to visually presented stop-signals that are similar in rhyme to the picture name than to those same stop-signals when they are phonologically dissimilar, but that it is no more difficult to stop in response to semantically similar stop-signals than to the same stop-signals when they are semantically dissimilar. A combined analysis of Experiments 3 and 4 using Experiment (i.e., stop-signal modality) as a between-subjects factor reveals significant main effects of similarity ($F1(1,94) = 7.50$, CI = ±2.1%; $F2(1,23) = 5.48$, CI = ±3.3%) and type of relationship ($F1(1,94) = 11.40$, CI = ±2.5%; $F2(1,23) = 10.34$, CI = ±3.8%), and an interaction between similarity and type of relationship (marginal by items; $F1(1,94) = 4.41$, CI = ±2.3%; $F2(1,23) = 4.10$, CI = ±3.6%, $p < .06$). As in the individual results of these experiments, planned comparisons revealed that this interaction results from a significant effect of similarity within the phonological condition ($F1(1,94) = 16.50$, $F2(1,11) = 14.75$), but no effect of similarity within the semantic condition ($F1(1,94) = 1.19$, $F2(1,11) < 1$). There was no main effect of Experiment and no factors interacted with Experiment, showing that the results did not differ based on stop-signal modality (auditory or visual). This, along with the results of the previous experiments, provides consistent evidence (as diagnosed by the current task) for a perceptual-loop-based monitor that is sensitive to phonological, but not semantic information.

An important conclusion that comes from the four experiments presented thus far is that the semantic relationship of a stop-signal to the to-be-produced word is unrelated to stopping performance. This suggests that a perceptual-loop-based monitor does not make a comprehension-to-production comparison at a level where semantic information is represented. Note that this runs counter to the standard assumption of how a perceptual-loop-based monitor might work (Levelt, 1983;1989;Levelt et al., 1999), and is particularly surprising in light of error-avoidance patterns that do clearly rely on some sort of semantic evaluation, for example the taboo words effect (Motley, et al., 1982).

It is, of course, possible that the reason a semantic effect did not emerge is because the stop-signal task employed in these experiments does not adequately simulate normal self-monitoring processes. The stop-signal task might only require speakers to make a peripheral comparison between phonological word forms (rather than processing the stop-signal and comparing via the perceptual loop), in which case the lack of semantic effects in these experiments might not reflect perceptual loop monitoring but might instead reflect demands specific to this task. There is good evidence that self-monitoring is sensitive to at least some types of semantic information (e.g., Motley et al., 1982), so it is perhaps surprising that this task has so consistently failed to find evidence for a semantically based comparison.

One way to test whether the current experimental paradigm is sensitive to perceptual-loop functioning more broadly rather than just involving a phonological discrimination task is to determine whether known non-phonological influences on monitoring affects performance in this task. For example, just as speakers are particularly adept at detecting and halting the production of errors that would result in taboo words (Motley et al., 1982), they should find it especially easy to detect and halt speech in response to stop-signals that are taboo words. However, if the experimental paradigm simply implements a kind of phonological detection task, then there should be no difference between stopping performance to taboo and to neutral words. Furthermore, if the semantic relationship of the stop-signal to the picture name did not affect stopping performance in the previous experiments simply because speakers were focusing only on the more peripheral phonological information from the stop-signal word, then the inclusion of emotionally charged words might encourage speakers to semantically process the stop-signals (possible evidence of which would come from showing sensitivity to emotionally charged words), and thus might reveal any effect of semantic similarity.

There is another potential procedural reason that might have obscured an effect of the semantic relationship of the stop-signal to the picture name in the previous experiments. Specifically, the stop-signal may have simply been presented too late to allow semantic effects to emerge. When comprehending a word, semantic information is likely to be available somewhat later than phonological information (especially for auditorily presented words). In the tasks in all these experiments, a floor SOA of about 300 ms (or a fixed SOA of 300 ms in Experiment 4) was imposed. Thus the stop-signal could only have been processed for about 300 ms (assuming it takes about 600 ms to name a picture) before the picture-name was articulated. In this vein, it is worth noting that a numerical difference was observed in Experiment 3 between semantically similar and dissimilar conditions; it is possible that this nonsignificant difference would be more robust if speakers had time to process the stop signals to the level of semantics before committing to articulating the target names.

Some evidence that may be consistent with this idea comes from a study that used a similar task to the one used in this experiment (Levelt, et al., 1991). In their study, subjects named pictures and, on a proportion of the trials, were also presented with an auditory test probe slightly after the picture presentation to which they had to provide a lexical decision. In one experiment (Experiment 3) the auditory probes were either nonwords or words that were either identical, semantically or phonologically related, or unrelated to the picture name. Interestingly, in their medium-SOA condition (the SOAs of their test probes varied by items, but the average SOA for their medium-SOA condition was 373 ms), they found that subjects were slower to provide a lexical decision to a word phonologically related to the picture name than to one semantically related or unrelated to the picture name. However, in the short-SOA condition (averaging 73 ms), they found slower lexical decisions for both phonologically and semantically related words. These observations raise the possibility that semantic information about the stop-signal words in the present studies was simply not available early enough to be relevant to the stopping process. If so, then an effect of the semantic relationship of the stop-signal to the picture name should emerge when the stop-signal is presented at an earlier SOA.

Experiment 5

Experiment 5 addresses two possible problems with the experimental procedure used in the previous four experiments: the possibility that the stop-signal methodology as a whole is insensitive to semantic information, and the possibility that the timing constraints made semantic information available too late to be of use. Regarding the first possibility, if the failure to find effects of semantic relatedness in previous experiments was simply because the task is insensitive to semantic information in the stop-signals, then speakers should be similarly insensitive to taboo (or otherwise emotionally valent) stop-signals. In contrast, if speakers are better at halting speech in response to taboo or valent stop-signals than to neutral stop-signals, then the task is presumably sensitive to semantically based effects. If this is the case, the failure to find effects of semantic relatedness in the previous four experiments cannot be attributed to limitations of the experimental paradigm. Rather than use (sometimes highly offensive) taboo words, Experiment 5 tested this possibility with a weaker, but potentially still effective, manipulation of *valence* (or emotional charge). Specifically, speakers' ability to halt speech in response to valent stop-signal words (e.g., *cancer*) was compared to their ability to halt speech in response to neutral, but otherwise matched, control stop-signal words (e.g., *miller*).

The second possibility addressed by Experiment 5 is that the failure to find differences in stopping performance due to semantic relatedness of the stop-signal in the previous experiments was because semantic information was not available early enough to be relevant. If this is the case, semantic effects should emerge when the speaker has more time to process the stop-signal before producing the picture name. On the other hand, if the semantic relationship of the stop-signal to the picture name truly does not affect stopping performance,

no differences should appear between the semantically related and unrelated conditions even at a much earlier SOA than used in the previous experiments. Experiment 5 tested this possibility by using the same procedure as Experiment 2, but with the stop-signals presented earlier – at a fixed SOA of 200 ms (an additional experiment tested an even earlier SOA of 150 ms, but found essentially the same results so will only be discussed briefly).

Presenting the stop-signals at this early of an SOA makes Experiment 5 temporally similar to picture-word interference tasks (e.g., Damian & Martin, 1999;Meyer & Schriefers, 1991;Rayner & Springer, 1986;Schriefers et al., 1990). Research using the picture-word interference paradigm has shown phonological facilitation for picture naming at an SOA of 200 ms: Pictures are named faster in conditions where a phonologically similar distractor appears shortly after the picture onset than in conditions where a phonologically dissimilar distractor appears (Starreveld, 2000). Thus the picture-naming process might be faster in the phonologically similar stop-signal conditions than in the phonologically dissimilar stop-signal conditions, which might, in turn, make it more difficult for subjects to halt word production since they would be further along in the production process. Because of this possibility, differences in stopping performance in the phonologically similar and dissimilar conditions should be interpreted with caution (as discussed below, this concern is less relevant to the previous four experiments).

The early SOA should not, however, pose a problem for comparing the semantically similar and dissimilar conditions, nor for comparing the emotionally valent and neutral conditions. Although studies using the picture-word interference paradigm show semantic-interference effects whereby pictures are named slower in conditions with semantically similar distractors than with semantically dissimilar distractors, semantic inhibition has generally been shown only when the distractor and picture are presented simultaneously or when the distractor word actually precedes the picture (although Bloem and colleagues have found semantic interference at a later SOA in a translation-task variant of the picture-word paradigm; Bloem & La Heij, 2003;Bloem, van den Boogaard, & La Heij, 2004). Therefore, any differences in stopping performance between the semantically similar and dissimilar conditions in this experiment can be assumed to reflect differences due to the semantic nature of the stop-signal, not due to differences in the time course of word production. And, as long as the emotionally valent and neutral words are semantically and phonologically unrelated to the picture name, they are unlikely to differentially affect speakers' picture naming latency.

Method

Participants—Forty-eight University of California, San Diego undergraduates participated in Experiment 5 in exchange for course credit. All participants reported English as their native language.

Materials, Design, and Procedure—The experimental design and stimuli were identical to those used in Experiments 1 and 2 (see Appendix A), with the addition of two conditions: the *emotionally valent* condition consisted of 18 words of relatively high valence (most of negative valence, e.g., *cancer*) and 18 control words (e.g., *miller*); the valent and neutral control stimuli are listed in Appendix C. The control words were of neutral valence and were individually matched to the valent words in length, lexical frequency, and mean bigram frequency using the English Lexicon Project web site (Balota et al., 2002). One control stimulus was inadvertently assigned to a picture to which it bore a semantic relationship (*toad – squirrel*), but because results were no different when this item was excluded, results are reported for all 18 items. The procedure was identical to that of Experiment 2, with visually presented stop and go-signals, with the exception of the SOA and SOA calibration. Instead of calibrating the SOA separately for each subject during the practice, the SOA was fixed at 200

ms for both the practice and the experimental trials, and subjects were given 72 practice trials and 432 experimental trials. Of the 432 experimental trials, 162 trials were go trials (each picture 9 times with no signal), 162 were go-signal trials (each picture 9 times with the picture name as a go-signal), and 108 were stop-signal trials (each picture 6 times, once with each type of stop-signal).

Results

Because stopping accuracy in Experiment 5 was relatively good (presumably because of the earlier SOA), results are reported using the criteria that speakers must have completely avoided production of the picture name to count as a successful stop. However, unlike in the previous experiments, different criteria for classifying successfully stopped trials led to quantitatively different results in Experiment 5. These differences will be discussed where relevant.

Figure 8 shows the mean stopping accuracy as a function of phonological and semantic similarity of the stop-signal to the picture name in Experiment 5. The pattern of results is similar to that found in the previous four experiments, but with greater stopping accuracy overall. In particular, subjects successfully stopped less often in the phonologically similar conditions (47.9% of stop-signal trials) than in the phonologically dissimilar conditions (60.9% of stop-signal trails), but subjects appeared to have no more difficulty stopping in the semantically similar conditions than in the semantically dissimilar conditions (54.4% of stop-signal trials in each case). Statistical analyses confirm these observations, with a significant main effect of phonological similarity by subjects ($F1(1,47) = 76.8$, CI = ±3.0%) and by items ($F2(1,17) = 71.8$, CI = ±3.3%) but no significant effect of semantic similarity by subjects or by items ($F1(1,47) < 1$, CI = ±2.2%; $F2(1,17) < 1$, CI = ±2.8%). The interaction between phonological and semantic similarity was marginally significant by subjects ($F1(1,47) = 4.02$, CI = ±3.7%), but not significant by items ($F2(1,17) = 2.12$, CI = ±4.9%). (In an additional experiment with an even shorter SOA of 150 ms, there was no significant interaction between phonological and semantic relatedness, suggesting that this marginal interaction is not robust.)

Figure 9 shows the mean stopping accuracy as a function of the emotional valence of the stop-signal words. Speakers more successfully halted speech in response to emotionally valent stop-signals (65.2% of stop-signal trials) than to matched control stop-signals (58.8% of stop-signal trials), although this difference was notably smaller when using the two-phoneme criteria for successfully stopped trials (a 2.2% difference). Statistical analyses support these observations. Stopping performance was significantly better to the emotionally valent stop-signals then to their controls ($F1(1,47) = 8.57$, CI = ±4.4%; $F2(1,17) = 11.8$, CI = ±3.7%, though this comparison was not significant when using the two-phoneme criteria for successful stops; $F1(1,47) = 1.05$, CI = ±4.2%; $F2(1,17) = 1.09$, CI = ±4.0%).

Expressed in terms of SSRTs, this stopping performance corresponds to slower SSRTs to phonologically similar than to phonologically dissimilar stop-signals (373 ms vs. 336 ms; $F1(1,42) = 18.5$, CI = ± 17.2 ms; $F2(1,17) = 48.4$, CI = ± 9.29 ms), faster SSRTs to emotionally valent than to neutral stop-signals (323 ms vs. 344 ms; $F1(1,42) = 5.67$, CI = ±18.2 ms; $F2(1,17) = 11.0$, CI = ±7.26 ms), but no difference in SSRTs to semantically similar and dissimilar stop-signals (355 ms vs. 354 ms; $F1(1,42) < 1$, CI = ± 11.8 ms; $F2(1,17) < 1$, CI = ± 6.20 ms). Note that five subjects were excluded from both SSRT analyses by subjects because of missing values occurring when all responses in a condition were successfully stopped.

The picture-naming latencies for Experiment 5 are reported in Table 3. Eight subjects were excluded from the RT analysis of phonological by semantic relatedness, and 10 subjects excluded from the RT analysis of valence, because of missing values occurring when all responses in a condition were successfully stopped or due to equipment malfunction. Surprisingly, speakers were significantly *slower* to respond in the phonologically related

conditions than in the phonologically unrelated conditions (568 ms vs. 545 ms; $F1(1,39) = 7.82$, CI = ±16.7 ms; $F2(1,17) = 5.69$, CI = ±22.0 ms). Response times were not significantly different in the semantically related and semantically unrelated conditions (553 ms vs. 561 ms; $F1(1,39) < 1$, CI = ±16.7 ms; $F2(1,17) = 1.01$, CI = ±13.4 ms), and there was no significant interaction between phonological and semantic relatedness ($F1(1,39) = 2.30$, CI = ±21.1 ms; $F2(1,17) < 1$, CI = ±24.0 ms). There was also no difference between response times in the emotionally valent and control conditions ($F1(1,37) < 1$, CI = ±32.5 ms; $F2(1,17) < 1$, CI = ±32.6 ms).

Discussion

Experiment 5 found that subjects were better able to stop speech in response to emotionally valent stop-signals than to neutral, but otherwise matched, stop-signal words. Thus, subjects must have processed the stop signals semantically. Nevertheless, subjects stopped speech about equally in the semantically similar and dissimilar stop-signal conditions, even with a very short SOA. Together, these observations suggest that the failure to find effects of semantic similarity in the previous four experiments is unlikely to be due to subjects simply not semantically processing the stop-signals nor to the length of the delay between the picture onset and the onset of the stop-signal word (and further suggests that the small difference in the semantic condition of Experiment 3 was spurious).

The finding that subjects were better able to stop when presented with emotionally valent words shows that this task can capture patterns of data that require semantic evaluation of an error. The fact that the effect emerged even with a relatively weak manipulation of valence indicates that the effect might be even stronger with the sorts of taboo words used in the classic demonstration by Motley et al. (1982). Thus the stop-signal task is likely to capture speakers' ability to detect and halt production of taboo words, despite showing no suggestion that monitoring is influenced by semantic similarity. This suggests that the taboo words effect may not be due to a perceptual loop monitor making a semantically based comparison, but rather to emotionally valent and taboo words capturing attention (see, e.g., Pratto & John, 1991;MacKay et al., 2004).

Experiment 5 also replicates the effect of phonological similarity found in Experiments 1, 2, and 4 (and the marginal effect found in Experiment 3), such that subjects had a harder time stopping when presented with a phonologically similar stop-signal than with a phonologically dissimilar stop-signal. Although this might be expected to result from phonological facilitation of the picture-naming process at this short of an SOA, when subjects did fail to stop they were actually slower in the phonologically related conditions. Nevertheless, it may be that phonological facilitation, and not monitor function, underlies at least part of the differences in stopping performance between the phonologically similar and dissimilar conditions in Experiment 5 (note that reaction times were assessed on a different set of trials than stopping performance), as a larger numerical difference was observed between the phonologically similar and dissimilar conditions in Experiment 5 [13%] than in Experiment 2 [5%].

General Discussion

Summarizing, Experiment 1 showed that subjects had a harder time stopping in response to auditorily presented stop signals that were phonologically similar (onset-related) to picture names than to stop signals that were phonologically dissimilar, whereas they showed no difference in stopping performance to stop signals that were semantically similar to picture names as compared to stop signals that were dissimilar. Experiment 2 used visual stop-signals and Experiments 3 and 4 used rhyme-related stop-signals (with auditory and visual presentation, respectively) and found similar results (albeit with smaller differences). This suggests that the effects are not due to the later arrival of discrepant information in the

phonologically similar conditions because the same pattern emerges both with visually presented and with rhyme-similar stop-signals. Experiment 5 showed that subjects were better able to stop speech in response to emotionally valent stop-signals than to neutral controls, and otherwise found the same pattern of stopping performance as in the previous four experiments, despite using earlier presentation of the stop-signals. This suggests that the lack of semantic effects was not due to speakers failing to process stop signals semantically for either task-demand or time-course reasons.

These experiments show that when speakers compare their to-be-produced speech to simultaneously comprehended speech, they are vulnerable to any phonological similarity between the two, whereas they do not seem to be vulnerable to semantic similarity between the two. The implications of these results are important primarily for the perceptual-loop theory of speech monitoring, and are also important for other situations where speakers' productions are modified by simultaneously comprehended language, including when speakers are subjected to changes in auditory feedback and when speakers are engaged in dialogue. These issues are discussed in turn.

To begin, the standard perceptual-loop theory of speech monitoring (Levelt, 1983,1989) assumes that speakers compare to-be-produced speech to (internally) comprehended speech at the level of concepts or semantics. If so, then in general, speakers should have had difficulty determining that to-be-produced words were different from simultaneously comprehended words that were conceptually similar. These experiments repeatedly failed to find any evidence to support this prediction. Similarly, according to the standard perceptual-loop theory, speakers should be insensitive to the phonological relationship between to-be-produced words and simultaneously comprehended words. Contrary to this prediction, speakers in these experiments did have difficulty determining that to-be-produced words were different from simultaneously comprehended words that were phonologically similar.

These observations suggests that speakers compare to-be-produced words to simultaneously comprehended words at the level of phonology, implying that within the framework of the perceptual loop, speakers might similarly monitor their to-be-produced language at the level of phonology. This fits with corpus-based evidence that internal monitoring is sensitive to phonological similarity (Nooteboom, 2005), and with the claim that the monitor is worse at detecting errors in the comprehension cohort of the target word (Roelofs, 2004). The latter observation might also explain why the difference between phonologically related and unrelated stop-signals was greater for onset-related stop-signals (Experiments 1, 2, and 5) than for rhyme-related stop-signals (Experiments 3 and 4), as the effect of rhyming competitors is typically weaker than the effect of onset related (cohort) competitors (see, e.g., Allopenna, Magnusson, & Tanenhaus, 1998).[2]

Although a phonologically based monitor is contrary to the standard idea of how a perceptual-loop monitor works (namely that the monitor makes a semantic or conceptual level comparison), a monitor that makes comparisons at a level of phonology might be functionally advantageous for a number of reasons. One is that a phonological-level monitor would presumably work faster than a conceptual level monitor, because the comprehension system would have to do less processing before comparing formulated to intended output (on the assumption that comprehending to the level of phonology can be completed more quickly than comprehending to the level of semantics). A second reason concerns the problem of the comparison itself, and is the complement of the 'common-code' observation noted in the introduction. That is, it was noted above that a perceptual-loop monitor might compare to-be-produced to comprehended expressions at the level of meaning or concepts, because conceptual

[2]Thanks to Rob Hartsuiker for pointing out this possibility.

representations are likely to be shared across the two linguistic modalities (which presumably makes the comparison process itself easier). However, the use of a representational level that is common between production and comprehension comes at a cost: Speakers must keep track of which representations are to-be-produced and which representations were (internally) comprehended. Such indexing or bookkeeping is likely to demand at least some attentional resources (an issue also raised by Vigliocco & Hartsuiker, 2002). In contrast, because articulation and acoustic analysis have distinct demands, it may be that phonologically based representations are distinct (or at least more easily distinguishable) between production and comprehension (e.g., Cutting, 1997;Levelt et al., 1999). Furthermore, current evidence suggests that even if such representations are distinct, they are tightly linked (e.g., Fowler, Brown, Sabadini, & Weihing, 2003), allowing for correspondence between representations at each level to be quickly assessed. Distinct but tightly coupled representations may reduce the need to keep track of which representations were to be produced and which representations were comprehended, making the comparison process itself less demanding. In short, a perceptual loop that detects erroneous performance by comparing a phonological (impending) output to internally comprehended input might be an efficient and workable mechanism.

Nonetheless, error detection must sometimes involve semantic comparisons, or speakers would be unable to detect errors that arose prior to phonological encoding. Furthermore, recent research has shown speakers to be sensitive to semantic but not to phonological similarity in a similar task – the opposite of the pattern presented here (Hartsuiker, Pickering, & De Jong, in press). In that task, speakers named pictures but had to halt their naming response on the small proportion of trials when the picture was replaced by another picture (300 ms after onset of the first) and name the second picture instead. Although Hartsuiker et al. were investigating the process of error repair rather than error detection, they found that speakers had more difficulty stopping the naming response to the first picture when the two pictures were semantically related than when they were unrelated (Experiment 1), but had no more difficulty halting when the pictures were phonologically related than when they were unrelated (Experiment 2). That is, Hartsuiker et al. (in press) found halting speech to be sensitive to semantic but not phonological information, whereas the experiments presented here found halting speech to be sensitive to phonological but not semantic information.

These contradictory findings suggest that speakers (and perhaps also the speech monitor) can halt speech based on the first comparison that is applicable: when phonological information is available first (as was the case with the word stop-signals in the experiments presented here), speakers make a comparison at a phonological level of representation, and when semantic information is available first (as was the case with the picture stop-signals in Hartsuiker et al., in press), speakers make a comparison at a semantic level of representation. In terms of perceptual loop monitoring, however, phonological information will always be available earlier than semantic information because the input to the monitor is the phonetic plan (or overt speech).

So how are speakers able to detect semantic errors that are phonologically well-formed? One possibility is that, in normal speech monitoring, speakers make comparisons both at a phonological and at a semantic level of representation. While this is not supported by the data reported here (nor by those reported by Hartsuiker et al., in press), some characteristics of the experimental paradigm might have encouraged speakers to rely solely on phonologically based comparisons for this task. This is unlikely to be due to insufficient processing of the stop-signals because speakers were better able to halt speech in response to emotionally valent words than neutral words (which must have required speakers to attend to semantic information – the emotional valence of a word is presumably part of its semantic representation). However, it is the case that phonological comparisons alone would be sufficient to detect differences between the stop-signals and the to-be-produced words. And while it is not clear why speakers would

have so overwhelmingly eschewed semantic comparisons in this task, as these would have been equally sufficient for the task demands and presumably would have lead to better performance (especially on phonologically similar stop-signal trials), we cannot rule out the possibility that speakers could have made semantic level comparisons but simply chose not to.

An alternative possibility is that the detection of phonological errors relies on the perceptual loop (and thus exhibits similarity-based vulnerability), but the detection of semantic errors relies on other mechanisms. This proposal fits with other evidence showing distinctions between the detection of phonological and semantic errors (e.g., Nootboom, 2005; Postma & Noordanus, 1996) and suggests that a phonologically sensitive perceptual-loop monitor leads to only some of the error patterns that speech-monitoring is generally used to explain. For example the mixed-error effect described above – that erroneously produced words are more similar both semantically and phonologically to the intended words they replace than is expected by chance – can be explained by claiming that a phonological-level perceptual-loop monitor is less likely to halt production of semantic-word substitution errors that are also phonologically similar, compared to semantic-word substitution errors that are phonologically distinct.

In contrast, detection of semantic errors may not rely on the perceptual loop, but rather involve a judgment of whether the word fits semantically in the relevant phrase. If this is the case, then speakers should be able to quickly detect and halt phonological errors via the perceptual loop even when naming single words and when under time pressure (as in these experiments), but may not be able to effectively monitor for semantic appropriateness (and perhaps also for other forms of linguistic orthodoxy, cf. Levelt, 1989) unless under conditions where they can inspect whole phrases. Similarly, while a phonologically based monitor is inconsistent with a perceptual loop based explanation of the taboo words effect, Experiment 5 showed that that speakers were better able to halt production in response to emotionally valent words despite showing no effect of semantic similarity. This suggests that the taboo words effect might be better explained by the attention capturing nature of taboo and emotional stimuli (an explanation similar to that of, for example, the *taboo Stroop effect*,Siegrist, 1995,MacKay et al., 2004, where naming the ink color of a word takes longer for taboo words than for neutral words), rather than as an effect of perceptual loop monitoring processes per se. By this account, the taboo words effect shows that we attend to our prearticulatory output, and that even internally generated taboo words can capture attention, thus avoiding erroneous articulation. Note that this explanation is not inconsistent with the perceptual loop explanation of the taboo words effect – by both accounts, taboo words are detected via the comprehension system – but the attention-grabbing characteristics of taboo words could allow detection even without a standard perceptual-loop comparison process. (Also note also that Levelt, 1989, assumes a third monitoring loop *within* conceptual processing – the appropriateness monitor – which presumably also helps speakers avoid erroneously producing taboo words in normal speech.)

A phonological-level monitor may also be unable to account for the lexical bias effect, because there is no clear reason why it would be particularly sensitive to lexical status. That is, a phonological comparison would not necessarily differentiate between words and nonwords because a nonword error could presumably be as phonologically similar to an intended utterance as a word error. This may not be a problem, however, if the lexical bias effect is not due to perceptual loop monitoring processes. In his corpus, Nooteboom (2005) found a lexical bias in speech errors, but also found that the correction rate (which presumably reflects overt repairing) for lexical and non-lexical errors was identical. This suggests either that different criteria apply to the internal and external perceptual loop monitoring channels (thus reducing the parsimony of the perceptual loop theory in general) or that the lexical bias is not due to perceptual loop monitoring processes at all. As mentioned in the introduction, others have proposed that the lexical bias effect might arise from feedback influences between phonological

and lexical levels of processing rather than from perceptual loop monitoring processes (Dell, 1986;Dell & Reich, 1981;Rapp & Goldrick, 2000;Humphreys, 2002), although evidence that strategic factors affect lexical bias complicate this account (Hartsuiker et al., 2005). As the current experiments did not include non-word stop-signals, the cause of the lexical bias effect is left for further research.

With respect to implications for speech monitoring, it is important to recognize that there are a number of limitations with this study. These experiments crucially rely on the assumption that speakers halt speech in response to discrepancies in comprehended self-produced speech (i.e., that the monitor is a perceptual-loop). Because the stop-signals were externally presented, the task model is only a simulation of speech monitoring if monitoring is carried out through the comprehension system. Furthermore, even if the monitor is best characterized as a perceptual loop, attentional processes may cause others' speech to be treated differently than self-generated speech. This would imply that the external stop-signals used in these experiments might not be the same as the comprehension of self-produced speech. In fact, this observation might explain why more consistent results were found in the experiments that presented the stop-signals visually (each showing about a 5% difference between phonologically similar and dissimilar conditions) than in the experiments that presented the stop-signals auditorily. Although both auditorily and visually presented stop-signals might be distinct from inner speech, the processing of the visual stop-signals might be less distinct from inner speech because the auditory basis of comparison may be internally generated. Thus visually presented stop-signals might make the task more analogous to the real perceptual loop.

It is also important to realize that there are several ways in which the task used here is a significant departure from the normal conditions of speaking and speech monitoring. Speakers named a small set of pictures multiple times, and were required to halt speech in response to "errors" (i.e., stop signals) relatively often, whereas normal speech is more varied, and (under most circumstances) less error filled. Stopping performance improved over the course of the task (significant correlations between trial order and stopping performance ranged from .40 to .58 in the five experiments), but additional analyses including experiment half as a factor (first vs. last half) showed no significant interactions between repetition and the factors of interest (with one exception: there was an interaction between phonological relatedness and experiment half in Experiment 5, significant by subjects but not by items, $F1(1,47) = 6.04$, CI = ±4.2%; $F2(1,17) = 1.68$, CI = ±5.7%; the lack of such an effect in the other four experiments suggests that this is not a robust finding.) Thus repetition does not appear to have influenced the main pattern of results, although because of the extensive practice sessions in these experiments, it is not possible to rule out the possibility that subjects would perform differently with less repetition and a lower proportion of stop-signal trials. Perhaps a more important concern is that while the underlying goal of speech monitoring is to achieve communicative success, these experiments do not require communication per se. Because of this, it is possible that under more realistic circumstances, speakers might make some other types of tests to determine if an error is disruptive enough to warrant interruption and repair.

Nonetheless, other considerations suggest that the observations from these experiments are yet important for the perceptual-loop theory of speech monitoring. The perceptual-loop theory is ubiquitous in language-processing accounts (e.g., Blackmer & Mitten, 1991; Levelt, 1983;Postma, 1997,2000;Postma & Kolk, 1993), including in the most comprehensive and influential of production theories in the field (Levelt, 1989;Levelt et al., 1999). Part of the reason for this ubiquity is that the perceptual loop, by virtue of being identified with the comprehension system, allows claims that speech monitoring is sensitive to everything that comprehension is sensitive to. Furthermore, because perceptual-loop influences have been claimed to be modulated by strategic factors or attention, any posited perceptual-loop influence can, within current accounts, be present or absent. The present experiments provide some

constraint on this powerful theory: Either the perceptual loop is primarily sensitive to phonological similarity between produced and simultaneously comprehended speech; or the perceptual loop employs mechanisms that allow it to distinguish internally- from externally-comprehended speech (an assumption that diminishes the otherwise desirable property of the perceptual-loop account that speakers can use the *same* mechanism to comprehend their own as well as others' speech); or monitoring does not occur with the comprehension system at all. Any of these implications constrain the perceptual-loop account more than it has been to date.

Furthermore, speakers' productions are sensitive to simultaneously comprehended language in situations other than those posited by the perceptual loop. Changes in speakers' perception of their own speech (e.g., by delayed auditory feedback) cause profound and involuntary disruption to production. The present results suggest that this might be because as speakers produce linguistic expressions, they are especially vulnerable to phonologically similar simultaneously comprehended language. Indeed, this fits with the observation that speakers' productions when under conditions of delayed auditory feedback are modified primarily in acoustic-articulatory ways (e.g., Elman, 1981;Houde & Jordan, 2002;Kalveram & Jancke, 1989), suggesting that delayed auditory feedback influences may also operate primarily at the level of phonological representation.

A second and very important circumstance under which speakers' productions are influenced by simultaneously comprehended language is during dialogue. Specifically, the *back-channel* signals that addressees convey to speakers can influence what speakers assume their addressees know, and therefore how their productions should continue (Clark & Krych, 2004;Yngve, 1970). The present results suggest that speakers may be able to evaluate at least some back-channel signals by comparing signals at phonological levels of representation. Thus, for example, the present account predicts that when addressees attempt to complete speakers' utterances (which is can be taken as a signal of addressees' successful comprehension), that speakers might be especially sensitive to such back-channels only when addressees use phonological content that speakers are expecting. For example, an addressee who says 'sofa' when a speaker was about to say 'sofa' might effectively convey that communication was successful, but an addressee who says 'couch' when the speaker intended 'sofa' may not as effectively convey that communication was successful.

Two other issues remain to be discussed – the reaction times and stop-signal reaction times. A potential concern is that some reliable differences exist between the picture naming latencies in the different stop-signal conditions of Experiments 1, 3, 4 and 5. Although these differences were not at all consistent across experiments, the fastest stop-signal conditions in Experiment 1 were also the ones with the worst stopping performance, raising the possibility of a kind of speed-accuracy trade-off. However, because the reaction times in Experiments 2, 3, 4, and 5 appear to be unrelated to the pattern of stopping performance, this kind of speed-accuracy trade-off cannot explain the entire pattern of effects. Furthermore, based on results from the picture-word interference literature (e.g., Cutting & Ferreira, 1999;Damian & Martin, 1999), any RT facilitation is likely to be relatively small and therefore may not be capable of driving all of the differences in stopping performance.

The stop signal reaction times (SSRTs) in these five experiments were considerably longer than what is usually reported in the stop-signal literature, even for stopping speech (e.g., Ladefoged, Silverstein, & Papcun, 1973, found that it takes about 200 ms to stop speech). This is not particularly surprising because traditional stop-signal tasks do not require as much processing of the stop-signal as the task used here, which requires linguistic processing to determine if the signal was a stop- or a go- signal. The estimates in these experiments, which average 299 ms across all five experiments, do fit relatively well with estimates of the time course of speech monitoring: Hartsuiker and Kolk (2001), drawing on a variety of sources,

estimate that the perceptual loop monitoring process takes approximately 350 ms (50 ms for audition, 100 ms for parsing, 50 ms for comparing, and 150 ms for interrupting). While this is slightly longer than the SSRT estimates in these experiments, if the estimates are modified for a model where the monitoring comparison process occurs at a phonological level of representation (as these data suggest may be the case) then the monitoring process might be expected to be somewhat faster since phonological representations might be available earlier than semantic ones. In this case, these SSRTs might match estimates of the time course of monitoring quite well.

Overall, these experiments suggest that as speakers produce language, they are especially sensitive to phonological relationships between what they are trying to say and what they simultaneously hear. This observation is important to consider in light of a perceptual-loop monitor, suggesting a phonological basis to its operation. It is also important to consider in terms of other production-comprehension relationships, including those arising with changes in auditory feedback and during dialogue. In general, speakers quickly and accurately produce utterances that successfully convey their meanings; an important component of that speed, accuracy, and success might be the tight and evidently phonologically-based relationship between the language they produce and the language they simultaneously comprehend.

Acknowledgements

Portions of this work were presented at the Architectures and Mechanisms for Language Processing conference in September, 2002. This research was supported by National Institute of Health grant R01 MH-64733. We thank Wind Cowles, Gary Dell, Zenzi Griffin and the members of the psycholinguistics group at UCSD for helpful discussions, Carla Firato, Cassie Gipson, Ana Molina, and Mary Tibbs for assistance collecting and transcribing data, and Jeff Bowers, Rob Hartsuiker, Sieb Nooteboom, and an anonymous reviewer for their comments on an earlier version of this paper. Please address correspondence to Bob Slevc at the Department of Psychology 0109, University of California, San Diego, La Jolla, CA, 92093-0109. Email: slevc@psy.ucsd.edu.

References

Allopenna PD, Magnuson JS, Tanenhaus MK. Tracking the time course of spoken word recognition using eye movements: Evidence for continuous mapping models. Journal of Memory & Language 1998;38:419–439.

Baars BJ, Motley MT, MacKay DG. Output editing for lexical status in artificially elicited slips of the tongue. Journal of Verbal Learning and Verbal Behavior 1975;14:382–391.

Baayen, RH.; Piepenbrock, R.; van Rijn, H. The CELEX lexical database [CD-ROM]. Philadelphia: Linguistic Data Consortium, University of Pennsylvania; 1993.

Balota, DA.; Cortese, MJ.; Hutchison, KA.; Neely, JH.; Nelson, D.; Simpson, GB.; Treiman, R. Washington University; 2002. The English Lexicon Project: A web-based repository of descriptive and behavioral measures for 40,481 English words and nonwords. http://elexicon.wustl.edu/

Blackmer ER, Mitton JL. Theories of monitoring and the timing of repairs in spontaneous speech. Cognition 1991;39:173–194. [PubMed: 1841032]

Bloem I, La Heij W. Semantic facilitation and semantic interference in word translation: Implications for models of lexical access in language production. Journal of Memory & Language 2003;48:468–488.

Bloem I, van den Boogaard S, La Heij W. Semantic facilitation and semantic interference in language production: Further evidence for the conceptual selection model of lexical access. Journal of Memory & Language 2004;51:307–323.

Caramazza A. How many levels of processing are there in lexical access? Cognitive Neuropsychology 1997;14:177–208.

Clark, HH. Using Language. New York, NY: Cambridge University Press; 1996.

Clark HH, Krych MA. Speaking while monitoring addressees for understanding. Journal of Memory & Language 2004;50:62–81.

Cohen JD, MacWhinney B, Flatt M, Provost J. PsyScope: A new graphic interactive environment for designing psychology experiments. Behavior Research Methods, Instruments & Computers 1993;25:257–271.

Coltheart M, Rastle K, Perry C, Langdon R, Ziegler J. DRC: A Dual Route Cascaded Model of Visual Word Recognition and Reading Aloud. Psychological Review 2001;108:204–256. [PubMed: 11212628]

Cutting, JC. The production and comprehension lexicons: What is shared and what is not. Unpublished doctoral dissertation, University of Illinois; Urbana-Champaign: 1997.

Cutting JC, Ferreira VS. Semantic and phonological information flow in the production lexicon. Journal of Experimental Psychology: Learning, Memory, & Cognition 1999;23:318–344.

Damian MF, Martin RC. Semantic and phonological codes interact in single word production. Journal of Experimental Psychology: Learning Memory, & Cognition 1999;25:345–361.

Dell GS. A spreading-activation theory of retrieval in sentence production. Psychological Review 1986;93:283–321. [PubMed: 3749399]

Dell GS, Reich PA. Stages in sentence production: An analysis of speech error data. Journal of Verbal Learning and Verbal Behavior 1981;20:611–629.

Dell, GS.; Repka, RJ. Errors in inner speech. In: Baars, BJ., editor. Experimental slips and human error: Exploring the architecture of volition. New York: Plenum; 1992. p. 237-262.

Elman JL. Effects of frequency-shifted feedback on the pitch of vocal productions. Journal of the Acoustical Society of America 1981;70:45–50. [PubMed: 7264071]

Fowler CA, Brown JM, Sabadini L, Weihing J. Rapid access to speech gestures in perception: Evidence from choice and simple response time tasks. Journal of Memory & Language 2003;49:394–413.

Fromkin VA. The non-anomalous nature of anomalous utterances. Language 1971;47:27–52.

Garrett, MF. The analysis of sentence production. In: Bower, GH., editor. The Psychology of Learning and Motivation. 9. New York: Academic Press; 1975. p. 133-177.

Goldrick M, Rapp B. A restricted interaction account (RIA) of spoken word production: The best of both worlds. Aphasiology 2002;16:20–55.

Houde JF, Jordan MI. Sensorimotor adaptation of speech I: Compensation and adaptation. Journal of Speech, Language, & Hearing Research 2002;45(2):295–310.

Hartsuiker RJ, Corley M, Martensen H. The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related beply to Baars et al. (1975). Journal of Memory & Language 2005;52:58–70.

Hartsuiker RJ, Kolk HHJ. Error monitoring in speech production: A computational test of the perceptual loop theory. Cognitive Psychology 2001;42:113–157. [PubMed: 11259106]

Hartsuiker RJ, Pickering MJ, De Jong NH. Semantic and phonological context effects in speech error repair. Journal of Experimental Psychology: Learning, Memory, and Cognition. in press

Humphreys, KR. Lexical Bias in Speech Errors. Unpublished doctoral dissertation, University of Illinois; Urbana-Champaign: 2002.

Kalveram KT, Jäncke L. Vowel duration and voice onset time for stressed and nonstressed syllables in stutters under delayed auditory feedback condition. Folia Phoniatr 1989;41:30–42.

Lackner, JR.; Tuller, BH. Role of efference monitoring in the detection of self-produced speech errors. In: Cooper, WE.; Walker, ECT., editors. Sentence processing. Hillsdale, NJ: Erlbaum; 1979. p. 281-294.

Ladefoged P, Silverstein R, Papcun G. Interruptibility of speech. Journal of the Acoustical Society of America 1973;54:1105–1108. [PubMed: 4757456]

Levelt WJM. Monitoring and self-repair in speech. Cognition 1983;14:41–104. [PubMed: 6685011]

Levelt, WJM. Speaking: From intention to articulation. Cambridge, MA: MIT Press; 1989.

Levelt WJM, Roelofs A, Meyer AS. A theory of lexical access in speech production. Behavioral & Brain Sciences 1999;22:1–75. [PubMed: 11301520]

Levelt WJM, Schriefers H, Vorberg D, Meyer AS, Pechmann T, Havinga J. The time course of lexical access in speech production: A study of picture naming. Psychological Review 1991;98:122–142.

Loftus GR, Masson MEJ. Using confidence intervals in within-subject designs. Psychonomic Bulletin & Review 1994;1:476–490.

Logan, GD. On the ability to inhibit thought and action: A users' guide to the stop signal paradigm. In: Dagenbach, D.; Carr, TH., editors. Inhibitory processes in attention, memory, and language. San Diego: Academic Press; 1994. p. 189-239.

Logan GD, Cowan WB. On the ability to inhibit thought and action: a theory of an act of control. Psychological Review 1984;91:295–327.

MacKay DG, Shafto M, Taylor JK, Marian DE, Abrams L, Dyer JR. Relations between emotion, memory, and attention: Evidence from taboo Stroop, lexical decision, and immediate memory tasks. Memory & Cognition 2004;32:474–488.

Maher LM, Rothi LJG, Heilman KM. Lack of error awareness in an aphasic patient with relatively preserved auditory comprehension. Brain & Language 1994;46:402–418. [PubMed: 7514943]

Marshall RC, Rappaport BZ, Garcia-Bunuel L. Self-monitoring behavior in a case of severe auditory adnosia with aphasia. Brain & Language 1985;24:297–313. [PubMed: 3978408]

Marshall RC, Tompkins CA. Verbal self-correction behaviors of fluent and nonfluent aphasic subjects. Brain & Language 1982;15:292–306. [PubMed: 7074346]

Marslen-Wilson WD, Welsh A. Processing interactions and lexical access during word recognition in continuous speech. Cognitive Psychology 1978;10:29–63.

Martin N, Gagnon DA, Schwartz MF, Dell GS, Saffran EM. Phonological facilitation of semantic errors in normal and aphasic speakers. Language & Cognitive Processes 1996;11:257–282.

Martin N, Weisberg RW, Saffran EM. Variables influencing the occurrence of naming errors: Implications for models of lexical retrieval. Journal of Memory & Language 1989;28:462–485.

Masson MEJ, Loftus GR. Using confidence intervals for graphically based data interpretation. Canadian Journal of Experimental Psychology 2003;57:203–220. [PubMed: 14596478]

McNamara P, Obler LK, Au R, Durso R, Albert ML. Speech monitoring skills in Alzheimer's disease, Parkinson's disease, and normal aging. Brain & Language 1992;42:38–51. [PubMed: 1547468]

Meyer AS, Schriefers H. Phonological facilitation in picture-word interference experiments: Effects of stimulus onset asynchrony and types of interfering stimuli. Journal of Experimental Psychology: Learning, Memory, & Cognition 1991;17:1146–1160.

Motley MT, Camden CT, Baars BJ. Covert formulation and editing of anomalies in speech production: Evidence from experimentally elicited slips of the tongue. Journal of Verbal Learning and Verbal Behavior 1982;21:578–594.

Nooteboom, SG. Listening to oneself: Monitoring speech production. In: Hartsuiker, RJ.; Bastiaanse, R.; Postma, A.; Wijnen, F., editors. phonological encoding and monitoring in normal and pathological speech. Hove, UK: Psychology Press; 2005. p. 167-186.

Oomen CCE, Postma A. Limitations in processing resources and speech monitoring. Language & Cognitive Processes 2002;17:163–184.

Osman A, Kornblum S, Meyer DE. The point of no return in choice reaction time: Controlled and ballistic stages of response preparation. Journal of Experimental Psychology: Human Perception & Performance 1986;12:243–258. [PubMed: 2943853]

Postma, A. On the mechanisms of speech monitoring. In: Hulstijn, W.; Peters, H.; Van Lieshout, P., editors. Speech production: motor control, brain research and fluency disorders. Amsterdam: Elsevier; 1997. p. 495-501.

Postma A. Detection of errors during speech production: a review of speech monitoring models. Cognition 2000;77:97–131. [PubMed: 10986364]

Postma A, Kolk HHJ. The covert repair hypothesis: Prearticulatory repair processes in normal and stuttered disfluencies. Journal of Speech and Hearing Research 1993;36:472–487. [PubMed: 8331905]

Postma A, Noordanus C. The production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. Language and Speech 1996;39:375–392.

Pratto F, John O. Automatic vigilance: The attention–grabbing power of negative social information. Journal of Personality and Social Psychology 1991;61:380–391. [PubMed: 1941510]

Rapp B, Goldrick M. Discreteness and interactivity in spoken word production. Psychological Review 2000;107:460–499. [PubMed: 10941277]

Rayner, K.; Pollatsek, A. The Psychology of Reading. Englewood Cliffs, NJ: Prentice Hall; 1989.

Rayner K, Springer CJ. Graphemic and semantic similarity effects in the picture -word interference task. British Journal of Psychology 1986;77:207–222. [PubMed: 3730727]

Roelofs A. Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). Psychological Review 2004;111:561–572. [PubMed: 15065924]

Schlenk K, Huber W, Wilmes K. Prepairs" and repairs: different monitoring functions in aphasic language production. Brain & Language 1987;30:226–244. [PubMed: 2436704]

Schriefers H, Meyer AS, Levelt WJ. Exploring the time course of lexical access in language production: Picture-word interference studies. Journal of Memory & Language 1990;29:86–102.

Siegrist M. Effects of taboo words on color-naming performance on a Stroop test. Perceptual & Motor Skills 1995;81:1119–1122. [PubMed: 8684902]

Snodgrass JG, Vanderwart M. A standardized set of 260 pictures: Norms for name agreement, image agreement, familiarity, and visual complexity. Journal of Experimental Psychology: Human Learning & Memory 1980;6:174–215. [PubMed: 7373248]

Starreveld PA. On the interpretation of onsets of auditory context effects in word production. Journal of Memory & Language 2000;42:497–525.

UCLA speech error corpus [Data file]. Linguistics Department, University of California; Los Angeles:

Vigliocco G, Hartsuiker RJ. The interplay of meaning, sound, and syntax in sentence production. Psychological Bulletin 2002;128:442–472. [PubMed: 12002697]

Wheeldon LR, Levelt WJM. Monitoring the time course of phonological encoding. Journal of Memory & Language 1995;34:311–334.

Wheeldon LR, Morgan JL. Phoneme monitoring in internal and external speech. Language & Cognitive Processes 2002;17:503–535.

Yates AJ. Delayed auditory feedback. Psychological Bulletin 1963;60:213–232. [PubMed: 14002534]

Yngve, VH. On getting a word in edgewise. Papers from the sixth regional meeting of the Chicago linguistic society; April 16–18, 1970; Chicago: University of Chicago Department of Linguistics. 1970. p. 567-578.

Appendix A

Stimuli Used in Experiment 1, 2, and 5 (adapted from Damian and Martin (1999), Experiment 3)

Picture name (go-signal)	Semantically similar stop-signal	Phonologically similar stop-signal	Semantically and phonologically similar stop-signal	Dissimilar stop-signal
apple	peach	apathy	apricot	couch
basket	crib	ban	bag	thirst
bee	spider	beacon	beetle	flag
bread	donut	brick	bran	nail
camel	pig	cash	calf	bucket
carrot	spinach	cast	cabbage	evening
duck	raven	dub	dove	brass
elephant	moose	elm	elk	stripe
fly	moth	flu	flea	rake
lamp	candle	landing	lantern	package
peanut	almond	piano	pecan	dress
rabbit	beaver	raft	rat	coffee
snake	eel	snack	snail	fire
spoon	ladle	sparkle	spatula	cable
squirrel	mole	skate	skunk	chain
train	bus	trophy	trolley	fox
truck	jeep	trap	tractor	celery
trumpet	horn	traffic	trombone	corner

Appendix B

Stimuli Used in Experiment 3 and 4

Picture name (go-signal)	Semantically similar stop-signal	Semantically dissimilar stop-signal	Phonologically similar stop-signal	Phonologically dissimilar stop-signal
basket	crib	elbow	casket	fling
bed	couch	cow	thread	chalk
bee	fly	cup	knee	yell
bell	gong	candle	yell	hole
book	journal	mop	cook	sand
bowl	cup	crib	hole	stamp
broom	mop	shirt	plume	cook
camel	giraffe	bus	mammal	fun
dog	fox	couch	log	knee
dress	shirt	whale	chess	prune
duck	swan	brooch	buck	heel
hand	elbow	fox	sand	mammal
lamp	candle	swan	stamp	casket
pear	orange	giraffe	wear	log
pig	cow	axe	wig	luck
ring	brooch	eel	fling	heap
saw	axe	goat	jaw	wear
seal	whale	ladle	heel	wig
sheep	goat	gong	heap	chess
snake	eel	journal	ache	jaw
sock	glove	planet	chalk	plume
spoon	ladle	orange	prune	buck
sun	planet	fly	fun	thread
truck	bus	glove	luck	Ache

Appendix C

Additional stimuli used in Experiment 5

Picture name (go-signal)	Emotionally valent stop-signal	Emotionally neutral stop-signal
apple	tsunami	summary
basket	murder	agreed
bee	polio	curio
bread	cancer	miller
camel	sex	add
carrot	death	field
duck	tornado	pivotal
elephant	bomb	mood
fly	scream	thread
lamp	deceit	tokens
peanut	horror	marble
rabbit	knife	looks
snake	famine	digest
spoon	danger	wonder
squirrel	doom	toad
train	gun	lot
truck	aids	beef
trumpet	disease	speaker

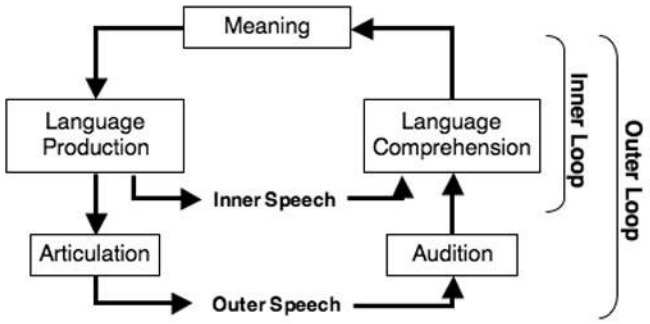

Figure 1.
The Perceptual Loop Theory of self-monitoring (Levelt, 1983,1989).

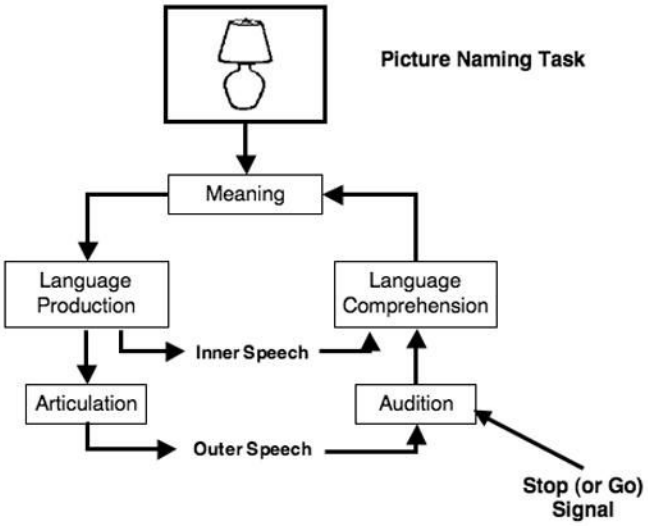

Figure 2.
Basic task model, mapped onto Levelt's (1983,1989) perceptual loop theory of speech monitoring. The drawing of a lamp represents the picture-naming task, and is assumed to initiate the word production process. The presentation of the stop-signal or go-signal is intended to "force feed" the comprehension system, mimicking the comprehension of self-produced speech (and thus of the inner monitor).

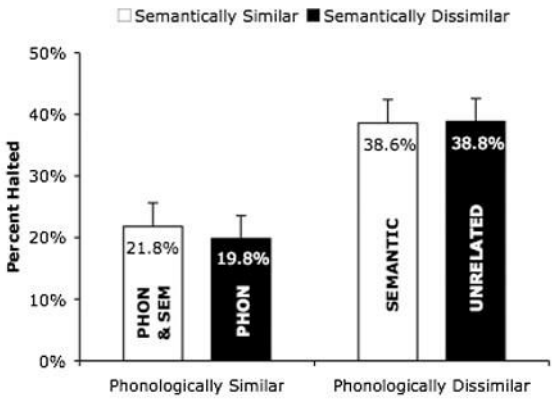

Figure 3.

Percentage of stop-signal trials successfully stopped as a function of phonological and semantic similarity of the stop-signal to the picture name in Experiment 1. Error bars indicate 95% confidence intervals. The “PHON” bar indicates stopping performance to stop signals phonologically similar to the picture name, the “SEMANTIC” bar indicates stopping performance to stop signals semantically similar to the picture name, the “PHON & SEM” bar indicates stopping performance to stop signals both phonologically and semantically similar to the picture name, and the “UNRELATED” bar indicates stopping performance to stop signals that are neither phonologically nor semantically similar to the picture name.

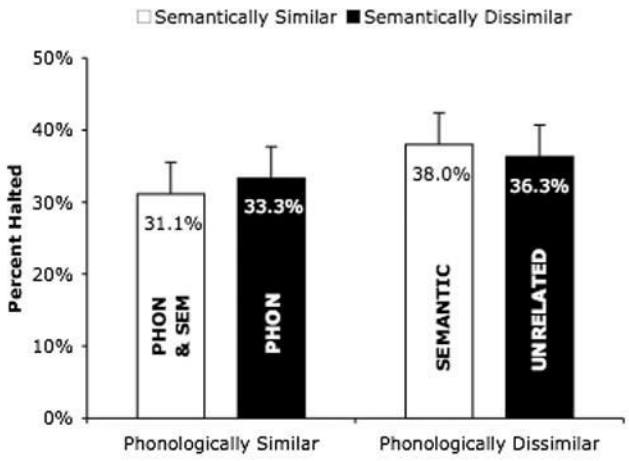

Figure 4.

Percentage of stop-signal trials successfully stopped as a function of phonological and semantic similarity of the stop-signal to the picture name in Experiment 2. Error bars indicate 95% confidence intervals. The “PHON” bar indicates stopping performance to stop signals phonologically similar to the picture name, the “SEMANTIC” bar indicates stopping performance to stop signals semantically similar to the picture name, the “PHON & SEM” bar indicates stopping performance to stop signals both phonologically and semantically similar to the picture name, and the “UNRELATED” bar indicates stopping performance to stop signals that are neither phonologically nor semantically similar to the picture name.

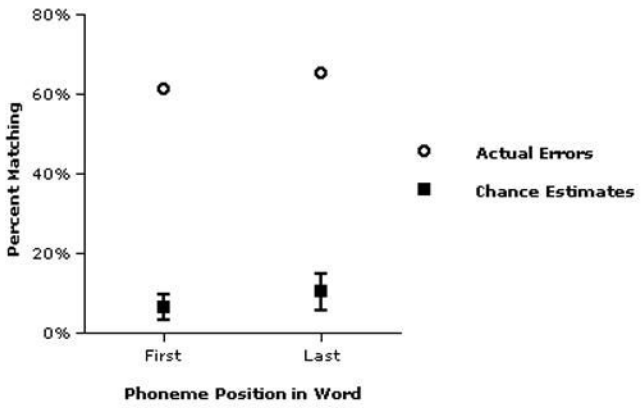

Figure 5.

Percentage of phonemes that match in the first and last position of actual and intended words from 274 word substitution errors (Malapropisms) in the UCLA Speech Error Corpus (open circles) and from chance estimates derived form 100 re-pairings of the actual and intended words of the same errors (filled squares). The bars around the filled squares indicate the range of the 100 chance estimates.

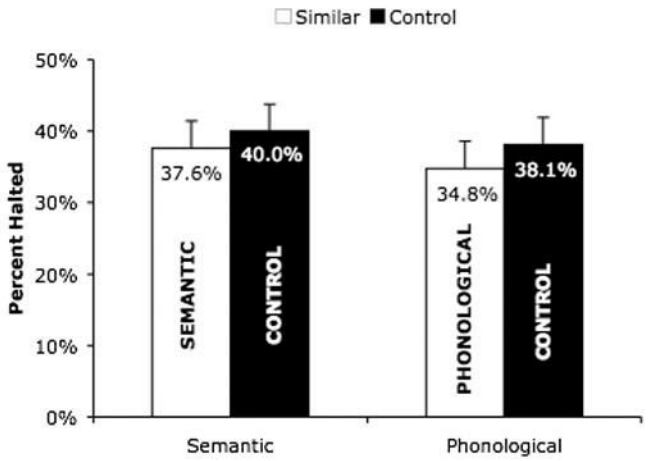

Figure 6.

Percentage of stop-signal trials successfully stopped as a function of similarity (similar or dissimilar) and type-of-relationship (phonological or semantic) of the stop-signal to the picture name in Experiment 3. Error bars indicate 95% confidence intervals. The “SEMANTIC” bar indicates stopping performance to stop signals that are semantically similar to the picture name, the “PHONOLOGICAL” bar indicates stopping performance to stop signals that are phonologically similar (i.e., that rhyme with) the picture name, and the “CONTROL” bars indicate stopping performance to these same stop-signals when paired with pictures to which they are neither phonologically nor semantically similar.

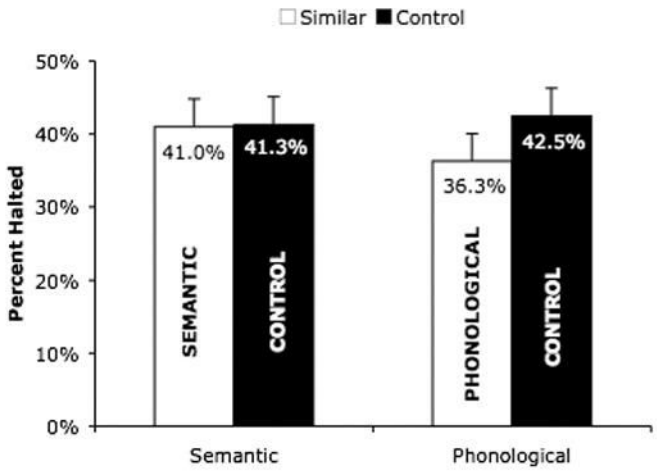

Figure 7.

Percentage of stop-signal trials successfully stopped as a function of similarity (similar or dissimilar) and type-of-relationship (phonological or semantic) of the stop-signal to the picture name in Experiment 4. Error bars indicate 95% confidence intervals. The “SEMANTIC” bar indicates stopping performance to stop signals that are semantically similar to the picture name, the “PHONOLOGICAL” bar indicates stopping performance to stop signals that are phonologically similar (i.e., that rhyme with) the picture name, and the “CONTROL” bars indicate stopping performance to these same stop-signals when paired with pictures to which they are neither phonologically nor semantically similar.

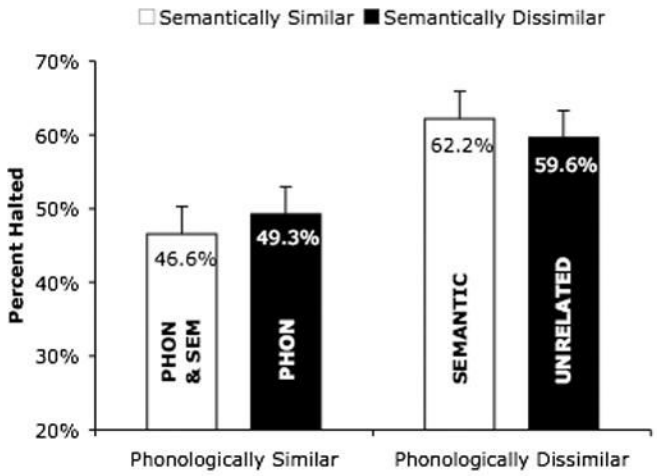

Figure 8.

Percentage of stop-signal trials successfully stopped as a function of phonological and semantic similarity of the stop-signal to the picture name in Experiment 5. Error bars indicate 95% confidence intervals. The “PHON” bar indicates stopping performance to stop signals phonologically similar to the picture name, the “SEMANTIC” bar indicates stopping performance to stop signals semantically similar to the picture name, the “PHON & SEM” bar indicates stopping performance to stop signals both phonologically and semantically similar to the picture name, and the “UNRELATED” bar indicates stopping performance to stop signals that are neither phonologically nor semantically similar to the picture name. (Note that the scale is different than in the previous figures.)

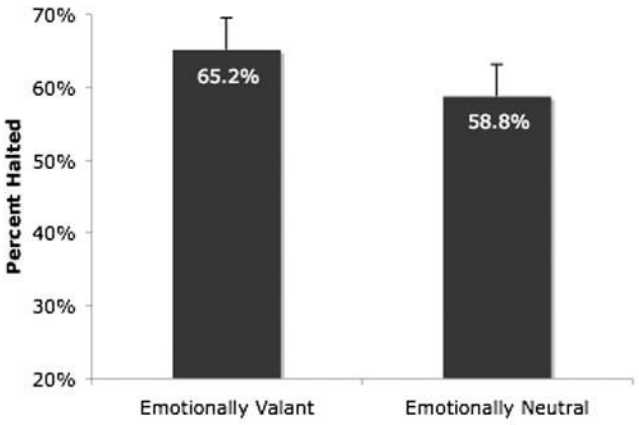

Figure 9.

Percentage of stop-signal trials successfully stopped as a function of the emotional valence of the stop-signal words. Error bars indicate 95% confidence intervals. The left bar indicates stopping performance to emotionally valent words (e.g., *cancer*) and the right bar indicates stopping performance to matched control words of neutral valence (e.g., *miller*).

Table 1

Mean Reaction Times for Experiments 1 and 2 by Type of Trial

Trial Type	Experiment 1	Experiment 2
Go	569	572
Go-signal	555	579
Stop - Phon. & Sem.	532	552
Stop - Phon.	536	555
Stop - Sem.	548	554
Stop - Dissimilar	561	547

Note. Reaction times are all measured in milliseconds; “Phon. & Sem.” indicates stop-signals both phonologically and semantically similar to the picture name, “Phon.” indicates phonologically similar stop-signals, “Sem.” indicates semantically similar stop-signals, and “Dissimilar” indicates stop-signals dissimilar to the picture name.

Table 2

Mean Reaction Times for Experiments 3 and 4 by Type of Trial

Trial Type	Experiment 3	Experiment 4
Go	537	584
Go-signal	524	593
Stop - Phon.	518	543
Stop - Dissimilar Phon.	524	543
Stop - Sem.	514	530
Stop - Dissimlar Sem.	524	533

Note. Reaction times are all measured in milliseconds; “Phon.” indicates stop-signals phonologically similar (in rhyme) to the picture name, “Dissimilar Phon.” indicates phonologically dissimilar stop-signals, “Sem.” indicates semantically similar stop-signals, and “Dissimilar Sem.” indicates semantically dissimilar stop-signals.

Table 3

Mean Reaction Times in Experiment 5 by Type of Trial

Trial Type	Experiment 5
Go	567
Go-signal	574
Stop - Phon. & Sem.	558
Stop - Phon.	577
Stop - Sem.	547
Stop - Dissimilar	544
Stop - Valent	555
Stop - Neutral	560

Note. Reaction times are all measured in milliseconds; “Phon. & Sem.” indicates stop-signals both phonologically and semantically similar to the picture name, “Phon.” indicates phonologically similar stop-signals, “Sem.” indicates semantically similar stop-signals, “Dissimilar” indicates stop-signals dissimilar to the picture name, “Valent” indicates emotionally valent stop-signals, and “Neutral” indicates emotionally neutral stop-signals.