

Motor movement matters: The flexible abstractness of inner speech

GARY M. OPPENHEIM AND GARY S. DELL
University of Illinois at Urbana-Champaign, Urbana, Illinois

Inner speech is typically characterized as either the activation of abstract linguistic representations or a detailed articulatory simulation that lacks only the production of sound. We present a study of the speech errors that occur during the inner recitation of tongue-twister-like phrases. Two forms of inner speech were tested: inner speech without articulatory movements and articulated (mouthed) inner speech. Although mouthing one's inner speech could reasonably be assumed to require more articulatory planning, prominent theories assume that such planning should not affect the experience of inner speech and, consequently, the errors that are "heard" during its production. The errors occurring in articulated inner speech exhibited the phonemic similarity effect and the lexical bias effect—two speech-error phenomena that, in overt speech, have been localized to an articulatory-feature-processing level and a lexical-phonological level, respectively. In contrast, errors in unarticulated inner speech did not exhibit the phonemic similarity effect—just the lexical bias effect. The results are interpreted as support for a flexible abstraction account of inner speech. This conclusion has ramifications for the embodiment of language and speech and for the theories of speech production.

One of the earliest ideas about thinking was that it is nothing more than inner speech, a weakened form of overt speech in which movements of the articulators occur but are too small to produce sound (Watson, 1913). A remarkable experiment by S. M. Smith, Brown, Toman, and Goodman (1947) demonstrated that this idea was false. Abolishing any trace of articulation through curare-induced total paralysis (requiring a respirator!) did not impair the participant's (Smith, himself) ability to think or understand his colleagues' speech. So Watson's bold claim (i.e., thought = inner speech = articulation) could not be true.

Nevertheless, there remains a great deal of interest in inner speech and the role of motoric processes in language and cognition. In this article, we report an experiment that (without curare) elucidates the relationship between articulation and inner speech imagery.

Many recent studies have investigated the extent to which linguistic representations have sensory or motoric components—the central question of the embodied-cognition framework. Embodiment, in the domain of language processing, is usually taken to be about whether meaning is sensory-motor in nature—specifically, in terms of engaging sensory or motor simulations of the events signified by linguistic referents (e.g., Barsalou, 1999; Lakoff, 1987; Pulvermüller, 2005). For instance, understanding the word *reach* may require basic visual, auditory, proprioceptive, and motoric circuitry to simulate the act of reaching so that the main difference between actually reaching and merely understanding the word *reach* is an apparent lack of motor movement.

But a second question arises concerning embodiment and language: Speech, regardless of its meaning, is the result of motor action. Given this, do internal representations of speech have motor components? The motor theory of speech perception (Galantucci, Fowler, & Turvey, 2006; Liberman, Delattre, & Cooper, 1952; Liberman & Mattingly, 1985) is a classic example of a theoretical stand on this question. It claims that listeners interpret an acoustic speech signal as the result of specific articulatory movements, covertly simulating the movements as a step toward recognizing the semantic conditions that created them. Thus, covert articulatory simulations are posited to play a central role in the mapping from sound to meaning. A second example, and the one that we investigate here, concerns the nature of inner speech—the silent, internal speech that Watson (1913) referred to. Although we know that inner speech is not the basis of thought, it does accompany and clearly supports many cognitive activities, such as planning, reading, and memorization (e.g., Sokolov, 1972; Vygotskiĭ, 1965; see also, e.g., J. D. Smith, Reisberg, & Wilson, 1992, for how auditory verbal imagery contrasts with inner speech).

Inner speech is generally thought of as the product of a truncated overt speech production process. Theories differ, however, about where this truncation lies (see Figure 1). Of particular interest is the distinction between a phonological level of processing (with units called *phonemes* or *phonological segments*) and an articulatory level (with units called articulatory or phonetic *features*). Phonemes, in the speech production process, are sometimes consid-

G. M. Oppenheim, goppenh2@illinois.edu

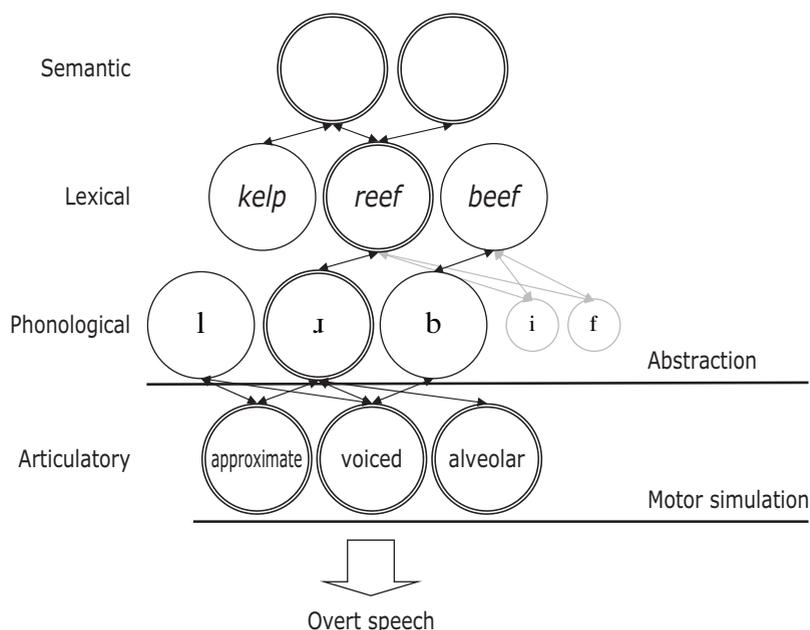


Figure 1. Theories of inner speech offer competing claims for its locus of generation, illustrated within the framework of Dell's (1986) model of speech production.

ered the lowest level at which abstract linguistic representations are actively constructed (e.g., Dell, 1986; Levelt, Roelofs, & Meyer, 1999). Phonemes are ultimately realized as articulatory behavior that is subject to constraints on articulator positions (e.g., /b/: lips together), but they do not by themselves specify the individual motor components that achieve the target positions (e.g., raise the jaw) in the context of neighboring phonemes. So if inner speech represents an embodied form of imagery, in the sense of incorporating sensory-motor simulations, it might be expected to go beyond the level of the phoneme and specify some articulatory features.

According to one class of theories, the motor simulation view (e.g., Dell, 1978; Levelt, 1989; Postma & Noordanus, 1996; cf. Reisberg, Smith, Baxter, & Sonenshine, 1989), inner speech is much like overt speech: It is fully specified for details like articulatory features. Compared with overt speech, it merely lacks observable sound and movement. For instance, in Levelt's (1983, 1989) original perceptual monitoring theory, inner speech emerges when an articulatory plan is internally monitored by the speech comprehension system. Monitoring this inner speech during overt speech production then allows a speaker to intercept errors before they would normally be heard. So the key point in motor simulation theories is an assumption that inner speech necessarily includes articulatory detail.

A second class of theories—the abstraction view—holds that inner speech is the consequence of the activation of abstract linguistic representations (e.g., Caplan, Rochon, & Waters, 1992; Dell & Repka, 1992; Indefrey & Levelt, 2004; Levelt et al., 1999; MacKay, 1992; Oppenheim & Dell, 2008; Vygotskiĭ, 1962; Wheeldon & Levelt, 1995). The crucial difference here is that inner speech emerges before a speaker retrieves any articula-

tory information. For example, Wheeldon and Levelt described a form of inner speech that corresponds to Levelt et al.'s¹ phonological word form level: syllabified strings of phonemes. Since Levelt et al.'s model used only feed-forward activation, these phonemes cannot reflect lower level information like articulatory features. Therefore, in the abstraction view, it is hypothesized that the experience of inner speech (i.e., speech production imagery) should be quite unconcerned with actual motor simulations. As we will review below, much evidence supports the abstraction view.

Six Arguments for the Abstractness of Inner Speech

Inner speech is faster than overt speech. Although the durations for tasks involving inner and overt recitation of particular words are highly correlated, overt speech tasks take longer, suggesting that inner speech production is abbreviated in some manner (e.g., MacKay, 1981, 1992). Although this finding does not require that inner speech specifically stop at the phoneme level, it does suggest that inner speech lacks a full specification of articulatory properties.

Inner speech uses fewer natural resources than does overt speech. Similarly, although neuroimaging studies have shown that inner speech production activates many of the same brain areas as overt speech (e.g., Paulesu, Frith, & Frackowiak, 1993; Yetkin et al., 1995), these areas tend to be less active in inner speech, suggesting that processing in these areas is not as complete or reliable as in overt speech. Specifically, inner speech involves less activation of brain areas thought to subserve the planning and implementation of motor movements (e.g., Barch et al., 1999; Palmer et al., 2001; Shuster &

Lemieux, 2005; see also McGuire et al., 1996, and Shergill et al., 2001, for neuroimaging work that differentiates between inner speech and auditory verbal imagery). These physiological observations certainly do not compel an abstraction view, since less motor activation may still be sufficient to produce motor simulations. But to the extent that activation of motor and premotor areas corresponds to the psycholinguistic concepts of articulatory encoding or simulation, the reduced activation for inner speech suggests that these processes are weak or incomplete.

Inner speech does not require articulatory abilities. The ability to overtly articulate a word is not required for successful use of inner speech. Anarthric patients, who have brain lesions that disrupt overt articulation (e.g., BA 1, 2, 3, 4, or 6), nevertheless show indirect signs of intact inner speech (Baddeley & Wilson, 1985; Vallar & Cappa, 1987). For instance, Baddeley and Wilson's patient G.B., who could not speak at all, showed normal accuracy when assessing pairs of written nonword homophones—a task that should have involved producing the strings in inner speech. Similarly, in a modern analogue to the curare experiment, localized magnetic interference (i.e., rTMS) can disrupt healthy participants' overt speech while leaving their inner speech seemingly intact (Aziz-Zadeh, Cattaneo, Rochat, & Rizzolatti, 2005). These dissociations suggest that inner speech can perform some of its functions without the contribution of articulation-specific processes.

Articulatory suppression does not (necessarily) eliminate inner speech. A motor simulation view of inner speech requires that its production engages articulatory resources, implying that articulatory suppression (e.g., in the form of concurrent articulation tasks, such as repeating "tah tah tah") should profoundly impair performance on tasks that require inner speech. One such task is phoneme monitoring (e.g., is there a /v/ sound in the word *of*?), in which a word's pronunciation must be internally generated and monitored for the presence of a target phoneme. But phoneme-monitoring experiments have not consistently demonstrated the predicted impairments. Although J. D. Smith et al. (1992) reported that articulatory suppression impaired phoneme-monitoring accuracy for orthographically presented words, Wheeldon and Levelt (1995) demonstrated only a very limited cost from suppression when participants monitored their internally generated Dutch translations of auditorily presented English words for phoneme targets. Therefore, although speakers may employ fine-grained articulatory simulations under some circumstances (e.g., when processing orthographic stimuli in J. D. Smith et al., 1992), these may not be required for all inner speech tasks.

Inner speech practice does not (necessarily) transfer to overt speech performance. If inner and overt speech involve comparable planning processes, one should expect that practicing an utterance in inner speech would improve overt performance and vice versa. Such transfer definitely occurs when the purpose of the internal practice is to rehearse conceptual or lexical information. For instance, MacKay (1981) had bilingual participants

practice a sentence in one language (German or English), using either inner or overt speech, before testing their production of the same sentence in a second language. Inner and overt practice yielded equivalent improvements in subsequent overt performance. A more everyday example: Mentally rehearsing a shopping list is quite effective for later reproduction of the list. Inner speech practice is less effective, however, when the overt speech task is articulatorily challenging. Dell and Repka (1992) found that although overtly practicing a tongue twister improved the accuracy of both inner and overt performance, inner practice improved only inner performance. This asymmetry suggests that inner speech production may fail to engage the articulatory specification that contributes to tongue twisters' ability to twist tongues. Together, the studies suggest that inner speech involves relatively intact higher level representations (e.g., on the word or message level), but incomplete lower level representations (e.g., articulatory features).

Phonological errors in inner speech do not show a phonemic similarity effect. Oppenheim and Dell (2008) compared self-reported speech errors made during the overt or inner recitation of tongue twisters. Phonological errors in both conditions tended to produce more words than nonwords (i.e., lexical bias; e.g., *reef* → *leaf* is more likely than *wreath* → *leath*), which, according to Dell's (1986) spreading activation model, suggests intact lexical-phonological processes in both inner and overt speech. But only the overt speech errors tended to involve similarly articulated phonemes (i.e., the phonemic similarity effect; e.g., *reef* → *leaf*, where /r/ and /l/ are both articulated as voiced approximants, is more likely than *reef* → *beef*, where /r/ and /b/ share only the voiced feature). It is well known that speech errors are strongly affected by shared features (e.g., Goldrick, 2004; MacKay, 1970), and hence, the absence of an effect of shared features on the inner speech errors suggests that inner speech involves phonemic but not subphonemic (e.g., articulatory) representations. This result is therefore consistent with the distinction presented in Figure 1 between a processing level concerned with lexical-phonological representations (present in inner speech) and a postphonological level at which featural information is relevant (present in overt speech). Also note that such a distinction has been well supported by studies of errors by individuals with brain damage. For example, Goldrick and Rapp's (2007) patient BON's errors suggested an articulatory planning deficit, whereas their patient CSS had difficulty retrieving phonological segments.

Challenges for the Abstraction View (i.e., Support for the Motor Simulation View of Inner Speech)

However, the abstraction view cannot offer a full account of even those results that we have discussed above. For instance, the timing of inner and overt speech tasks ordinarily track each other with great regularity (e.g., silent sentence rehearsal, MacKay, 1981, 1992; reading, Abramson & Goldinger, 1997), suggesting that inner speech preserves aspects of the temporal details of overt speech.

Moreover, the fact that inner speech activates brain areas involved in motor planning (e.g., Barch et al., 1999; Yetkin et al., 1995) suggests at least some degree of motor simulation. In fact, for decades, cognitive neuroscientists have been sufficiently convinced of the correspondence between inner and overt speech that they regularly used inner speech as a proxy for overt speech in neuroimaging tasks (e.g., Sahin, Pinker, Cash, Schomer, & Halgren, 2009; Thompson-Schill, D'Esposito, Aguirre, & Farah, 1997).

Furthermore, although Oppenheim and Dell's (2008) main finding—that inner slips are insensitive to phonemic similarity—is suggestive, it contradicts the conclusions of some other inner speech error studies. For instance, both Dell (1978) and Postma and Noordanus (1996) compared inner and overt speech performance on tongue twisters and concluded that the two conditions elicit similar types of errors in remarkably similar distributions (cf. Meringer & Meyer, 1895, cited in MacKay, 1992). Moreover, Postma and Noordanus found that participants reported similar numbers and types of errors (e.g., phoneme anticipations, perseverations) when reciting tongue twisters in inner speech, mouthed speech (i.e., inner speech with silent articulatory movements), and noise-masked overt speech (i.e., overt speech without auditory feedback), with higher reporting rates only in the normal overt speech condition. This pattern supports the motor simulation view, suggesting that unarticulated inner speech and normal overt speech engage similar planning processes—right down to the level of individual motor movements—with any apparent differences being attributable to auditory error detection. In addition, the specific claim of Oppenheim and Dell (2008) that inner speech does not involve featural information is, itself, controversial. For instance, Brocklehurst and Corley (2009) found that phonemic similarity promoted phoneme slips in inner speech—an unlikely result if inner speech is never specified at a subphonemic level. Our study, reported here, will be directly relevant to this controversy.

Finally, as was mentioned earlier, inner speech is thought to play a role in overt speech production (e.g., Levelt, 1983). By monitoring an inner speech version of an utterance before overtly articulating it, speakers can avoid, interrupt, and possibly correct slips even before they finish articulating them. Analyses of such quickly interrupted speech errors (e.g., *reef* → *le . . .*; e.g., Nooteboom & Quené, 2008), which must have been interrupted before articulatory movements were initiated (Levelt, 1983), demonstrate that some amount of detailed articulatory information must be available prior to overt speech. Specifically, these analyses suggest that errors in which a target phoneme is replaced by one that shares fewer articulatory features (e.g., /r/ → /b/) are, under some conditions, more likely to be quickly interrupted than those in which the target and replacement share more features (e.g., /r/ → /l/). If these errors are in fact detected by monitoring inner speech, inner speech would have to specify the articulatory features in order for its monitoring to produce similarity effects for interrupted overt errors. In our study, we also address this apparent inconsistency.

Synthesis: A Flexible Abstraction Account of Inner Speech

Perhaps a shortcoming of both the abstraction and motor simulation views lies in conceiving of inner speech as a stable, consistent phenomenon. We have reviewed many previous findings suggesting that speakers can produce a fairly abstract type of inner speech, as well as the sort of fully realized overt speech that produces observable sound and movement. Therefore, speakers seem to have the ability to control the extent of motoric expression in speech production. Given this control, we ought to be able to vary the extent to which inner speech is motoric. In terms of Figure 1, we hypothesize that activation could be restricted to the level of phonemes in one situation (activating articulatory features weakly, inconsistently, or not at all) but that articulatory features could be strongly activated in another.

Here is the general framework. We follow Levelt and colleagues' more recent characterization that inner speech emerges on a phonological level of representation (e.g., Indefrey & Levelt, 2004; Levelt, 1999; Wheeldon & Levelt, 1995). In their model and in Dell's (1986) spreading activation model, word forms are actively selected at the phoneme level. We therefore specifically associate the creation of inner speech with the act of phoneme selection. Note that the cascading activation specified in Dell's (1986) model implies that phoneme selection may actually occur *after* articulatory features have already been activated. Feedback from those features can therefore bias the phoneme selection process, thereby affecting inner speech. Specifically, we predict that increasing the amount of activation that articulatory features get in inner speech should elicit phonemic similarity effects on inner slips, and we suggest that variation in this activation can explain the variability in the extent to which we see articulatory effects in inner speech.

In our experiment, we investigated the question that concerned Oppenheim and Dell (2008) and Brocklehurst and Corley (2009): Are the slips that occur in inner speech sensitive to articulatory similarity? The previous research on this question has contrasted inner speech with overt speech, but inner and overt speech differ in two respects: sound and movement (Table 1). In the present study, we tested the three hypotheses concerning the abstraction or lack thereof in inner speech by comparing standard inner speech with an intermediate form: silent mouthed speech. Mouthed speech includes motor movements but generates no sound for a speaker to monitor. Therefore, we examined the slips made in two forms of silent speech: one that has an abundance of articulation (mouthed inner speech) and one that does not (unmouthed inner speech). In essence,

Table 1
Three Variants of Speech Production in Terms of Their Sensorimotor Properties

	Movement	Sound
Overt speech	+	+
Inner speech	–	–
Mouthed speech	+	–

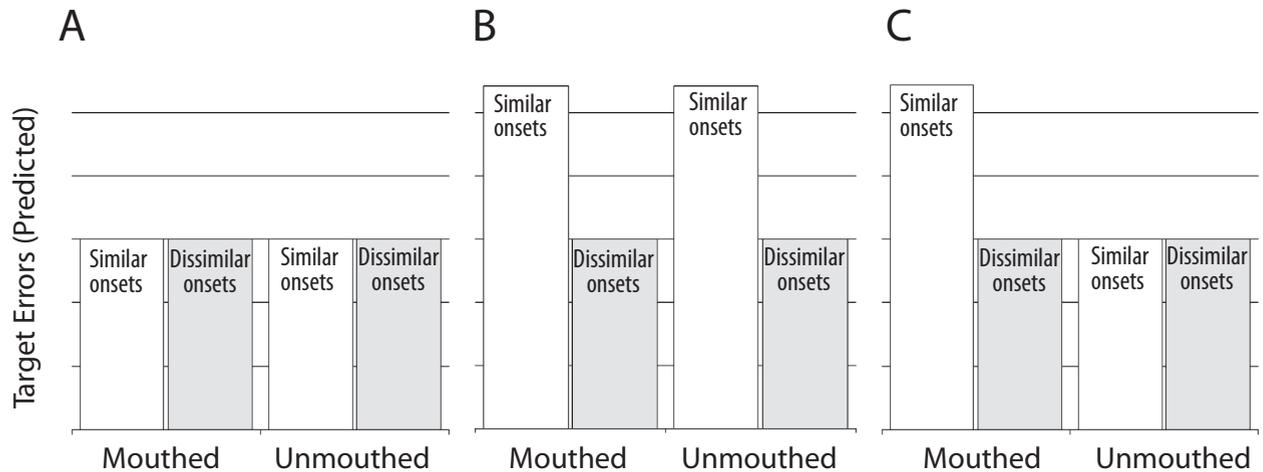


Figure 2. Three contrasting predictions for onset error patterns in mouthed and inner speech. (A) If inner speech necessarily consists of abstract linguistic representations (e.g., phonemes), then unmouthed inner speech errors should show no effect of articulatory similarity, and adding silent articulation should not change that. For instance, in Levelt's recent revision of his perceptual monitoring theory (see, e.g., Indefrey & Levelt, 2004), inner speech is specifically the product of a phonological encoding stage and, therefore, cannot incorporate information from subsequent phonetic encoding. Since mouthed speech creates no overt auditory signal and utterances can be monitored only by perceiving either inner speech or overt speech, silent mouthing has no opportunity to affect inner speech. (B) If inner speech necessarily consists of motoric simulations (see, e.g., activation at the level of articulatory features), then unmouthed inner speech should show a clear effect of articulatory similarity, and adding silent articulation should not make the effect any stronger. Although this prediction disagrees with our previous findings for inner speech, it is in keeping with other findings that we have reviewed (see, e.g., Brocklehurst & Corley, 2009; Postma & Noordanus, 1996). For instance, in his review of speech monitoring theories, Postma (2000) concluded that inner speech is phonetically specified, differing from overt speech only in terms of a monitorable acoustic signal. Since mouthed speech produces no such acoustic signal, its error reports should not differ from those of unmouthed inner speech. (C) Finally, if the experience of inner speech can incorporate additional articulatory planning, then mouthed speech should show a stronger effect of articulatory similarity than should unmouthed inner speech. To the extent that the mouthing requires a solid articulatory plan, the similarity effect there should be about as strong as that in overt speech.

the manipulation concerns the degree of articulation that participants include in their inner speech. Figure 2 presents the predictions of the three hypotheses.

Following Oppenheim and Dell (2008), in addition to examining how articulatory similarity affects phonological errors in inner speech, we also examined the effects of outcome lexicality (lexical bias). *Lexical bias* is the tendency of phonological speech errors to form words more often than nonwords. For example, a slip from /kæt/ (cat) to the word /bæt/ is more likely than a slip from /kæp/ to the nonword /bæp/. Lexical bias can be attributed to activation spreading between the phonological level and the lexical level (either automatically [e.g., Dell, 1986] or as the result of a monitoring process [e.g., Levelt et al., 1999]) and, therefore, indexes the activation of lexical and phonological information. The phonemic similarity effect, in contrast, derives from activation contacting a sub-phonemic-feature level by some means or other (e.g., Dell, 1986; Nooteboom & Quené, 2008). Our experimental manipulations thus targeted an abstract processing level and a more articulatory level.

METHOD

Participants recited tongue-twister phrases in which the opportunity for errors to exhibit both the phonemic similarity effect and the lexical bias effect was manipulated. In each trial, we used either silent unmouthed inner speech or silent mouthed speech. All conditions were manipulated in a within-participants and within-item-sets fashion for maximum power.

Participants

Eighty 18- to 30-year-old Champaign-Urbana residents participated in exchange for cash or course credit. All of them had normal or corrected-to-normal vision and hearing and were American English speakers who had not learned any other languages in the first 5 years of their lives.

Materials

Thirty-two matched sets of four-word tongue-twister phrases were devised, as is illustrated in Table 2. These were, for the most part,² the same stimuli used in Oppenheim and Dell (2008), designed to test for lexical bias on the third word and a phonemic similarity effect on the second and third words. The second word in each set was identical in all conditions, and the third word was identical within a condition of outcome lexicality.

Table 2
A Matched Set of Four-Word Tongue Twisters

	Similar Onsets				Dissimilar Onsets			
Word outcome	lean	reed	reef	leech	bean	reed	reef	beech
Nonword outcome	lean	reed	wreath	leech	bean	reed	wreath	beech

Phonemic similarity. We manipulated the similarity of onset consonants in terms of overlap in three major articulatory dimensions: place of articulation (e.g., /t/ vs. /k/), manner of articulation (e.g., /t/ vs. /s/), and voicing (e.g., /t/ vs. /d/).³ In the *similar* condition, the onset phonemes of the first and fourth words shared two features with the onsets of the second and third words. For example, in Table 2, the phonemes /r/ and /l/ are both voiced approximants but differ in place of articulation. In the *dissimilar* condition, the onsets of the first and fourth words (/b/) shared only one feature with the onsets of the second and third words. Referring to the Table 2 example again, /b/ and /r/ are both voiced but differ in their place and manner of articulation.

Lexical bias. Outcome lexicality was manipulated in the third word of each target set by a minimal change to its coda phoneme. In Table 2, for example, a slip replacing the /r/ onset of the target *reef* with an /l/ would create the word *leaf*. Changing the /f/ coda phoneme in *reef* to /θ/ (both /f/ and /θ/ are voiceless fricatives but differ in place of articulation) would create a new target word, *wreath*. And a slip, replacing the /r/ onset of the new target, *wreath*, with an /l/ would create the nonword *leath*. Since word frequency affects phonological errors (Dell, 1990; Kittredge, Dell, Verkuilen, & Schwartz, 2008), the third word of each set was controlled for target and slip-outcome log frequency (Kučera & Francis, 1967) [targets: lexical (*reef*), $M = 3.38$; nonlexical (*wreath*), $M = 3.28$; outcomes: lexical similar (*leaf*), $M = 2.73$; lexical dissimilar (*beef*), $M = 2.63$; nonlexical similar (*leath*), $M = 0.06$ ⁴; nonlexical dissimilar (*beeth*), $M = 0.0$].

These phrases were placed into counterbalanced lists, yielding four 32-item lists with eight phrases of each condition in each list. Within each list, half of the phrases in each condition were marked to be imagined while the participants mouthed them, and half were marked to be imagined without the participants' mouthing them; the order of these overtness conditions was pseudorandom and fixed. A second version of each of these four lists then reversed the mouthed/unmouthed pattern, resulting in a total of eight lists, with each participant assigned to one.

Procedure

Each trial consisted of an overt study phase followed by a silent testing phase. At the start of the study phase, a phrase appeared in the center of a 17-in. computer screen, in white 18-point Courier New font on a black background. After 3 sec, a quiet 1-Hz metronome began. The participants then recited the phrase aloud four times, in time with the metronome, pausing between repetitions, and then pressed the space bar to continue, signaling that they had memorized the phrase. The metronome then stopped, and the screen went blank for 200 msec.

Then the test phase began. A picture appeared in the center of the screen, cuing the participants to imagine saying the phrase while either mouthing (a picture of a mouth) or not mouthing (a picture of a head), and a faster (2-Hz) metronome began 500 msec later. The phrase reappeared in a small, low-contrast font at the top of the screen; the participants were instructed that they could check this between attempts or when reporting errors but that they should avoid looking at the words otherwise. The participants now attempted the phrase four times, pausing four beats between attempts. For the unmouthed condition, the participants were instructed to imagine saying the phrase in time with the metronome without moving their mouth, lips, throat, or tongue. The mouthed condition was identical, except that the participants were instructed to silently articulate while imagining the phrase. As they did so, the participants were instructed to monitor their inner speech in both conditions, stopping to report any errors immediately, and precisely specifying both their actual and intended utterances (e.g., "Oops, I said 'leaf reach' instead of 'reef leech'"). After reporting an error, they continued with their next attempt when they were ready. After completing the four fast attempts, the participants pressed the space bar, and the next trial began 200 msec later.

To ensure procedural consistency within and across participants, each participant was trained through two demonstration trials and

four practice trials (two mouthed and two unmouthed). In the rare case that a participant's reporting of an error was unclear, the experimenter prompted him or her for more information whenever it was possible to do so in a timely manner (e.g., participant: "Oh, I said 'leaf'." Experimenter: "'leaf' instead of what?" Participant: "I said 'leaf' instead of 'reef'"). The error reports were digitally recorded and transcribed both on- and offline.

Analyses

Any observed or self-reported deviations from the instructed procedures (e.g., mouthing in the unmouthed condition or reporting errors imprecisely, such as "Um, I said 'leaf'," or "Oops, I said . . . oh, nevermind.") were dealt with by excluding the affected trials (80 trials, <1%). Participants ($n = 17$) were replaced if this meant excluding more than 25% (i.e., four attempts) of their data from any one condition or more than 6.25% (i.e., eight attempts) of their data overall. Each of the remaining trials was categorized as follows, on the basis of self-reports. (1) *Target errors* were exactly those utterances in which one onset in the phrase replaced the other without creating other errors. For instance, replacing *reef* with *leaf* (a simple onset phoneme substitution) was considered a target error, whereas replacing *reef* with *lee*, *lead*, *leech*, *leafs*, or *leath* was not. (2) *Competing errors* included any errors, other than the target errors, in which the target onset could be construed as having been replaced by the other onset in the phrase. For instance, replacing *reef* with *lee*, *lead*, *leech*, *leafs*, or *leath* would all be considered competing errors. (3) *Other contextual word errors* were cases in which a participant reported misordering the second and third words or their codas (e.g., *reed* → *reef* or *reef* → *reed*). These are not directly relevant to our hypotheses, but they accounted for a substantial proportion of the nontarget errors. (4) *Miscellaneous errors* included all other errors, such as noncontextual onset errors, vowel or coda errors (e.g., *reef* → *rife*, *reef* → *real*), disfluencies (e.g., reported as, "I just stopped instead of saying the third word."), and other multiphoneme or word-level errors (e.g., *reef* → *rind*, *reef* → *frog*). (5) *No errors reported* trials were those trials on which no errors were reported.

Since the participants had been instructed to stop and report any errors immediately, in the rare event that a participant reported multiple errors in a single attempt, only the first of these errors was recorded.

Given the structure of the materials, our analyses were focused on just the second and third words of each item set. This strategy avoids potential confoundings arising from asymmetrical phoneme confusability (e.g., the probability of /s/ → /ʃ/ is greater than that of /ʃ/ → /s/; Stemberger, 1991) by holding the target onsets constant across all conditions. Thus, tests of phoneme similarity effects count errors on both the second and third words. Tests involving lexical bias, however, are restricted to target errors on the third word, which is the only part of the word set in which we systematically manipulated outcome lexicality (e.g., *reef* → *leaf*, *wreath* → *leath*, *reef* → *beef*, *wreath* → *beath*).

We computed the proportions of trials that contained target errors and report them with the raw error counts below. Any trials ending before the critical word was attempted—that is, before the second word in the similarity analyses and before the third word in the lexical bias analysis—were not included. Thus, converting the counts into proportions adjusts for the possibility that some utterances were interrupted before the critical words were attempted. Such opportunity-adjusted proportions were computed separately for each condition, participant (for the by-participants analyses), and item set (for the by-items consideration) and serve as the input for our statistical analyses.

We used Wilcoxon signed-rank tests, with a continuity correction (Sheskin, 2000), an adjustment for tied ranks (Hollander & Wolfe, 1973), and a reduction of the effective n when differences between paired observations were zero (e.g., Gibbons, 1985; Sheskin, 2000). Note that the Wilcoxon test is appropriate for any paired observations in which the size and sign of the differences are meaningful. This includes both main effect contrasts (dif-

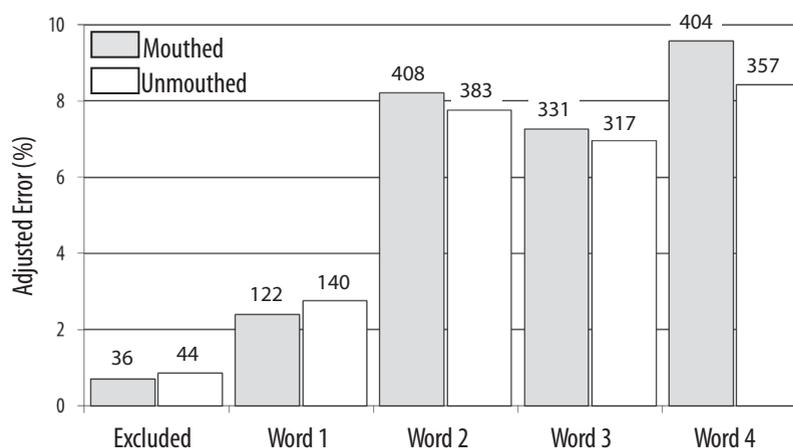


Figure 3. Total mouthed and unmouthed errors, collapsed across similarity and lexicality conditions, and plotted by position within the phrases. Each column shows the opportunity-adjusted error rate for that position, and the number above each column gives the unadjusted error count. Although excluded trials and errors on Words 1 and 4 were not included in our other statistical analyses, we report aggregate error counts here in order to demonstrate general similarities in terms of error distributions over word positions.

ference between two conditions) and interactions involving factors with only two levels (difference of differences). We rejected or failed to reject the null hypothesis at $\alpha = .05$ on the basis of the by-participants analyses, but to document the consistency of the effects across item groups for each contrast, we also examined the five item sets with the largest differences in either direction. Where null hypotheses were rejected, we report the number of those sets in which the difference was not in the overall direction (e.g., as “one out of five sets dissenting”). All planned tests of lexical bias, phoneme similarity effects, and similarity \times mouthing interactions are directional, on the basis of the findings of Oppenheim and Dell (2008) and related effects in the overt speech literature. We also report p values from directional tests when describing the outcome of tests of simple effects, even when our theoretical perspective favors null results (e.g., phonemic similarity in unmouthed inner speech; the implied direction is that errors are more likely in the similar condition).

RESULTS

The participants reported errors on roughly a quarter of the 10,240 recitations. Consistent with previous reports (e.g., Postma & Noordanus, 1996), this included approximately equal error totals for mouthed (1,301 errors) and unmouthed (1,241 errors) inner speech, suggesting that the mouthing manipulation did not greatly affect the overall probability of error production, detection, or reporting

in this experiment (main effect of mouthing, $p > .20$). Moreover, the errors showed similar distributions across word positions in the mouthed and unmouthed conditions (Figure 3). Such correspondence suggests that we can now interpret any differences in target error rates as reflecting underlying processes of inner speech production and comprehension or monitoring, rather than merely as differences in the base rates of error production. Furthermore, the closeness of the overall error counts suggests that mouthing does not particularly enhance the ability to detect errors. That is, we can agree with Postma and Noordanus’s conclusion that there is little contribution to error detection from any additional cues that articulation may provide.

Phonemic Similarity Effects

How does silent articulation affect the phonemic similarity effect? To address this question, we restrict our focus to the 366 targeted onset anticipations and perseverations on Words 2 and 3 (Table 3).

Here, we find that mouthed speech elicited a stronger phonemic similarity effect than did unmouthed inner speech (mouthing condition \times similarity interaction, $p < .03$, one out of five item sets dissenting). Target errors involving similarly articulated phonemes were signifi-

Table 3
Outcomes for Words 2 and 3, for Examination of Phonemic Similarity Effects

	Mouthed		Unmouthed	
	Similar	Dissimilar	Similar	Dissimilar
Target error	123 (0.050)	75 (0.030)	92 (0.037)	76 (0.031)
Competing error	39 (0.016)	29 (0.012)	30 (0.012)	27 (0.011)
Other contextual word error	120 (0.049)	135 (0.054)	115 (0.047)	118 (0.048)
Miscellaneous error	100 (0.041)	118 (0.047)	114 (0.046)	128 (0.052)
No error reported	2,087 (0.845)	2,132 (0.857)	2,106 (0.857)	2,126 (0.859)

Note—Counts for each outcome are followed, in parentheses, by the opportunity-adjusted proportion of trials that the count represents.

Table 4
Outcomes for Word 3, for Examination of Phonemic Similarity Effects

	Similar		Dissimilar	
	Word	Nonword	Word	Nonword
Mouthed				
Target error	49 (0.043)	24 (0.021)	31 (0.027)	16 (0.014)
Competing error	5 (0.004)	10 (0.009)	5 (0.004)	8 (0.007)
Other contextual word error	17 (0.015)	20 (0.018)	12 (0.011)	17 (0.015)
Miscellaneous error	33 (0.029)	25 (0.022)	38 (0.033)	21 (0.019)
No error reported	1,032 (0.909)	1,053 (0.930)	1,060 (0.925)	1,072 (0.945)
Unmouthed				
Target error	35 (0.03)	19 (0.017)	24 (0.021)	19 (0.017)
Competing error	3 (0.003)	10 (0.009)	8 (0.007)	6 (0.005)
Other contextual word error	13 (0.011)	10 (0.009)	21 (0.018)	11 (0.010)
Miscellaneous error	28 (0.024)	41 (0.037)	33 (0.029)	36 (0.032)
No error reported	1,071 (0.931)	1,034 (0.928)	1,064 (0.925)	1,062 (0.937)

Note—Counts for each outcome are followed, in parentheses, by the opportunity-adjusted proportion of trials that the count represents.

cantly more likely in the mouthed condition ($p < .002$, zero out of five sets dissenting) but not in the unmouthed condition ($p > .16$, three out of five sets dissenting). The lack of a significant phonemic similarity effect in unarticulated inner speech replicates Oppenheim and Dell's (2008) finding and reinforces the view that inner speech may fail to engage detailed articulatory representations, consistent with both the abstraction and the flexible abstraction views. And the strong similarity effect in mouthed speech is consistent with both the motor simulation and the flexible abstraction predictions. But the interaction, reflecting a stronger similarity effect in mouthed speech, is consistent only with the prediction that we derived for the flexible abstraction hypothesis. So although inner speech can operate on a more abstract form-based level, it can also incorporate lower level articulatory planning.

In addition, the presence of a similarity effect in mouthed speech also argues against an alternative explanation for the lack of a phonemic similarity effect found in unmouthed inner speech. One might imagine that errors involving similarly articulated phonemes are difficult to detect, because inner speech lacks sound (cf. Postma, 2000). Hence, errors in the similar phoneme condition would be underreported, and the overall effect of similarity diminished. In our mouthed speech condition, there was no sound, and yet the similarity effect was robust (5% similar slips, 3% dissimilar slips), roughly equivalent to what Oppenheim and Dell (2008) reported for an overt speech condition (4% similar slips, 2% dissimilar slips). Therefore, the presence of auditory information about the slip is not required for the similarity effect to be obtained. (Below, we test for and eliminate a more sophisticated form of this alternative, in which the similarity effect is attributed to a monitoring and repair process that works more effectively on mouthed slips.)

Lexical Bias

Both mouthed and unmouthed target errors showed significant lexical bias, consistent with the assertion that both types of inner speech engage higher level (e.g., phonemic

and lexical) representations (Table 4). Recall that outcome lexicality was specifically manipulated on Word 3 of each phrase. Target errors here produced more words than nonwords in both articulatory conditions (mouthed, $p < .006$, one out of five sets dissenting; unmouthed, $p < .03$, one out of five sets dissenting). No significant interaction between lexical bias and phonemic similarity (mouthed, $p > .36$; unmouthed, $p > .49$) or modality ($p > .22$) emerged. These results replicate previous findings of lexical bias for phonological errors in unarticulated inner speech (Brocklehurst & Corley, 2009; Oppenheim & Dell, 2008), extending them to articulated inner speech as well. Moreover, the lack of an interaction between articulation and lexicality contrasts with the presence of an interaction between articulation and phonemic similarity, reinforcing our interpretation that unarticulated inner speech is specifically impaired in terms of lower level articulatory representations (e.g., features).

Generation Versus Repair Accounts of Speech Error Effects

Until now, we have considered speech error patterns as being directly generated by the normal workings of the speech production process. This follows, for instance, suggestions that activation spreads between words and phonemes to create lexical bias and spreads between phonemes and articulatory features to create the phonemic similarity effect (e.g., Dell, 1986, 1990; Nozari & Dell, 2009). But an influential alternative—a whole class of explanations, actually—is that our speech error patterns may reflect error monitoring and repair biases. For example, biased repair accounts propose that speakers may initially generate word-outcome and nonword-outcome, or similar and dissimilar, errors in equal quantities but quickly catch and edit out errors of one kind more than errors of another kind (see also, e.g., Baars, Motley, & MacKay, 1975; Hartsuiker, 2006; Hartsuiker, Corley, & Martensen, 2005; Levelt et al., 1999; Nooteboom, 2005; Nooteboom & Quené, 2008). If so, our error effects could simply reflect the workings of a biased monitor/repair system and may not reveal anything about the core process of speech pro-

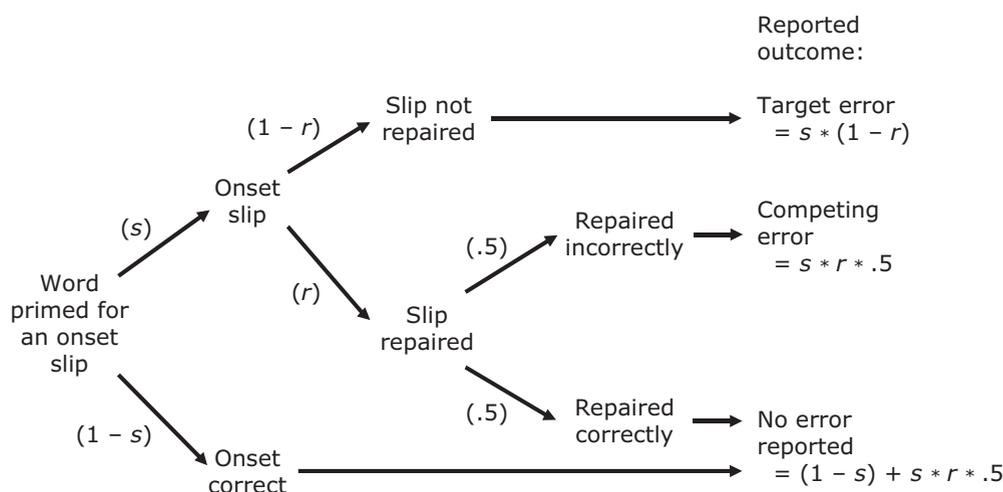


Figure 4. A multinomial processing tree describing possible roles for error generation and repair processes in creating the reported distributions of target errors, competing errors, and nonerror trials. The figures in parentheses give the probability of taking each tree branch. The equations at the right end specify the expected proportions of each outcome, in terms of the s and r parameters.

duction (e.g., Garnsey & Dell, 1984). Speech error effects that are already present in reports of inner speech errors, as we have demonstrated, would seem to rule out accounts in which monitoring is based on the perception of one's inner speech (e.g., Levelt 1983, 1989; Levelt et al., 1999). But other accounts propose a kind of fast and hidden monitoring mechanism (e.g., Nickels & Howard, 1995; Nootboom, 2005; Postma, 2000) by which errors may be aborted or repaired before speakers even become aware of them, suggesting that inner speech may already reflect monitoring biases. Therefore, the question of whether our error reports actually reflect speech production processes may require estimating the contributions of direct generation and biased repair mechanisms.

We can estimate these contributions by treating our data as the outcomes of a multinomial processing tree (e.g., Hu & Batchelder, 1994) in which slips are generated and repaired with varying probabilities (Figure 4; cf. Hartsuiker, Kolk, & Martensen, 2005). We follow Nootboom and Quené's (2008) insight that repair processes may manifest in errors that are repaired incorrectly, creating errors that would not otherwise occur in great quantities. For instance, a directly generated slip from *wreath* to *beeth* (the target error) may be internally repaired and ultimately reported⁵ as what we call a *competing error*, such as *beech*, *beet*, or *beesh*. For this exercise, we assumed that the prevalence of competing errors should index the repair process.

The resulting tree (Figure 4) assumes an error generation process followed by a monitoring/repair process. In the first stage, speakers select the phonemes of a word either correctly (e.g., *wreath* → *wreath*, with the probability $1 - s$) or as a target slip (e.g., *wreath* → *beeth*, with the probability s). Then all potential utterances are evaluated at the monitoring/repair stage. All correct utterances pass the monitor and are reported as correct. Of the initially generated target errors (s), the monitor

misses some (with probability $1 - r$), and the speaker reports them as target errors [yielding a total probability for target errors $s * (1 - r)$]. The monitor repairs the rest with probability r , yielding a total proportion of $s * r$ repaired utterances. These repairs produce either correct utterances ($s * r * .5$) or competing errors ($s * r * .5$) with equal⁶ probability [thus yielding total probabilities for correct utterances, $(1 - s) + (s * r * .5)$, and for competing errors, $(s * r * .5)$]. This assumed process specifies a system of equations by which we can estimate the underlying s and r parameters that might give rise to a particular distribution of target errors, competing errors, and correct responses. And by estimating repair biases (differences in r), we can focus on slip generation effects (differences in s).

The role of slip and repair biases in phonemic similarity effects. Following our reported interaction between similarity and articulation, we examined the mouthed and unmouthed conditions separately. For each condition, we first constructed a model in which the values of s and r are unrestricted. That is, they are allowed to vary as a function of similarity. These values were estimated from the error distributions for Words 2 and 3 (Table 3), using multiTree (Moshagen, 2010), and are given in Table 5. As was expected, the s value is numerically larger in the mouthed similar condition, echoing our major findings with the target errors. There is also a hint of an effect of similarity on r . To evaluate these numbers, we asked whether the s and r values necessarily differed between similarity conditions. Compared with the unrestricted model, using a single r parameter to account for both similar and dissimilar conditions did not noticeably worsen its account of the data in either mouthed speech [$G^2(1) = 0.48, p = .49$] or unmouthed inner speech [$G^2(1) = 1.29, p = .26$]. Hence, r is not significantly sensitive to similarity, and therefore, it does not appear that monitoring and repair biases played

Table 5
Estimates of Target Slip Generation and Repair Rates for an
Unrestricted Model, Derived From the Equations
Given in Figure 4 and the Data in Table 3

	Mouthed		Unmouthed	
	Similar	Dissimilar	Similar	Dissimilar
Target slip generation rate (<i>s</i>)	.0894	.0595	.0682	.0583
Error repair rate (<i>r</i>)	.388	.436	.395	.415

a major role in creating our phonemic similarity effects. What about *s*? Using the same *s* value for similar and dissimilar conditions does not hinder the model's account of unmouthed inner speech data [compared with the unrestricted model: $G^2(1) = 0.08, p = .78$]. In contrast, however, ignoring the similarity effect for *s* did offer a significantly worse fit for the mouthed speech results [$G^2(1) = 10.06, p < .002$]. Therefore, an effect of similarity in the mouthed condition can be attributed to the slip generation parameter. This is the same pattern that we reported earlier for target errors, but the multinomial analysis suggests that it still holds after taking possible monitoring and repair biases into account.

Such nested model comparisons do not offer an obvious way to test our phonemic similarity by articulation interaction, but by using an unrestricted model to derive separate parameter estimates for each participant, however, we were able to subject them to our standard Wilcoxon tests. The Wilcoxon analysis of the participant-derived parameters supported the conclusions from the nested model tests. For the *r* parameter, there was no significant similarity \times mouthing interaction for *r* ($p = .37$). The estimated slip rates (*s*), however, showed a significant interaction ($p < .04$), reflecting a stronger similarity effect for target slip generation in mouthed speech. Therefore, we conclude that our main empirical finding—that silent articulation restores the phonemic similarity effect—holds independently of influences of monitoring and repair biases as we have operationalized those biases here.

The role of slip and repair biases in lexical bias effects. Although they are encouraging, the results of the similarity analyses raise the question of whether our statistical methods were simply too weak to detect biased error repair rates. To address this question, we now turn to the lexical bias effect. From Baars et al. (1975) through Levelt et al. (1999), many researchers have argued for a role of monitoring biases in creating a lexical bias for overt speech errors. More recently, Hartsuiker, Corley, and Martensen (2005) and Nootboom and Quené (2008) suggested that lexical bias reflects the combined influences of slip generation and monitoring processes. So as a test case, we can examine whether analyzing our lexical bias data would reflect these contributions.

Again, we used our earlier analyses as a starting point. Since our earlier lexical bias analysis did not indicate interactions with either similarity or mouthing, we collapsed our data across those conditions. This gives us 217 target errors and 55 competing errors, split between word- and

nonword-outcome conditions, thus supplying close to the total number of competing errors (57) that we had analyzed for the similarity effect in unmouthed inner speech. The strategy, again, was to start with an unrestricted model, in which *s* and *r* were estimated separately for the word- and nonword-outcome conditions, and then to test whether using a single *s* or *r* parameter for the two conditions would significantly worsen the model's fit for the data.

Estimates from the unrestricted model are given in Table 6, indicating a slightly higher slip generation rate (*s*) for the word condition and a dramatically higher repair rate (*r*) in the nonword condition. Indeed, assuming a single repair rate (*r*) for both lexicality conditions clearly worsened the model's fit [$G^2(1) = 11.97, p < .0006$], suggesting that repair biases do contribute to the lexical bias effect. Here, the rates of competing errors show a large reverse lexical bias effect, leading to a conclusion that potential nonword outcomes are more likely to undergo repair. This is consistent with recent studies in which a role for monitoring and repair in lexical bias was posited (Hartsuiker, Corley, & Martensen, 2005; Nootboom & Quené, 2008). However, Nootboom and Quené and Hartsuiker et al. (2005) also described a role for slip generation (*s*) in creating the lexical bias effect, so we can ask whether our data reflect that contribution as well. The data here are less clear. The effect is in the predicted direction: *s* was greater for the word condition. But restricting the model to using the same *s* parameters while allowing the *r* parameters to vary freely produces only a marginal reduction in goodness of fit [$G^2(1) = 2.53, p = .11$]. So, the effect of lexicality on *s* is only marginal. In general, the multinomial analysis of lexical bias in our data supports recent conclusions in the overt speech literature, suggesting that the same processes that generate lexical bias in overt speech are active in inner speech. Moreover, the analysis makes us more confident about the similarity analysis, in which we found a substantial contribution of similarity to *s* in the mouthed condition only and little contribution of similarity to *r* in both the mouthed and unmouthed conditions.

Table 6
Estimates of Target Slip Generation and Repair Rates for an
Unrestricted Model, Derived From the Equations
Given in Figure 4 and the Data in Table 4

	Word	Nonword
Target slip generation rate (<i>s</i>)	.0413	.0337
Error repair rate (<i>r</i>)	.232	.466

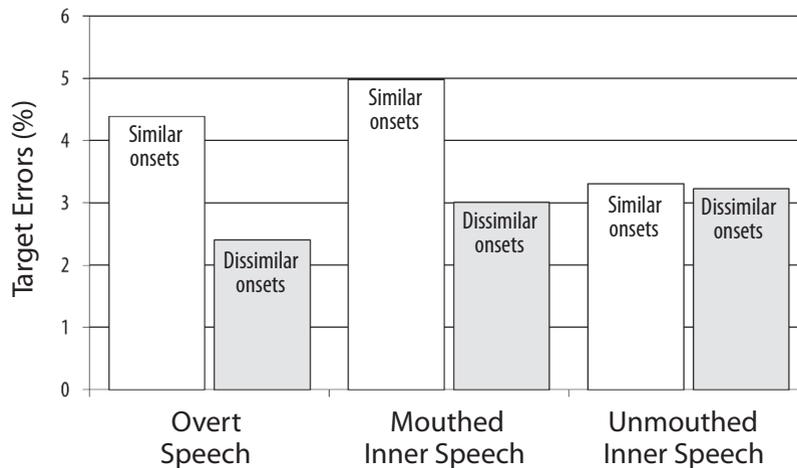


Figure 5. The big picture: Opportunity-adjusted target error rates for Words 2 and 3 in overt speech (Oppenheim & Dell, 2008), mouthed speech (present study), and unarticulated inner speech (weighted average from Oppenheim & Dell, 2008, and the present study).

DISCUSSION

How does articulatory information contribute to inner speech? Articulatory similarity matters—provided that the speech is articulated. This can be seen very clearly if we combine the present data with the overt and unmouthed inner conditions from Oppenheim and Dell (2008) (see Figure 5).

In the introduction, we outlined predictions for three proposals regarding the tendency of phonological errors to involve similarly articulated phonemes (the phonemic similarity effect): a motor simulation account, an abstract linguistic representation account, and a flexible abstraction account. The first two propose that inner speech either necessarily includes articulatory (phonetic) information or necessarily fails to include it, respectively. The last assumes some flexibility in this respect, and our major experimental finding—a robustly stronger effect of articulatory similarity in mouthed speech errors—requires such flexibility. Below, we briefly review the flexible abstraction hypothesis and the implications of our findings for theories of speech production.

The Flexible Abstraction Hypothesis, Elaborated

In a nutshell, our flexible abstraction account assumes that there is just one level for inner speech—a phonological level—but that it is affected by articulation. Assigning inner speech to a phonological level of representation is in accord with our previous work (Oppenheim & Dell, 2008), as well as with other models (e.g., Indefrey & Levelt, 2004; Roelofs, 2004; Wheeldon & Levelt, 1995). Accounting for the present data, however, requires some way for lower level information (e.g., articulatory features) to affect phonological representations. The models of Levelt and colleagues explicitly disallow such backward information flow. But two crucial properties of nondiscrete models (e.g., Dell, 1986)—cascading activation and

feature-to-phoneme feedback—allow the flexible abstraction hypothesis to account for the present data. Cascading activation means that features can be activated before a phoneme is selected, and the presence of feedback allows activated features to bias phoneme selection. Therefore, the phonemes that are selected can reflect varying amounts of activation beyond the phoneme level.

Speakers can, of course, control the extent of their overt articulation. They can yell, enunciate, murmur, whisper, mouth, or just say things in their heads, implying the ability to modulate the extent of articulation, and given that experiments in which articulatory similarity in unmouthed inner speech was examined have shown a nonsignificant reverse similarity effect (Oppenheim & Dell, 2008), a nonsignificant positive similarity effect (this article), and a significant similarity effect (Brocklehurst & Corley, 2009), it appears that articulatory activation in silent speech may also occupy intermediate values.

Our description so far has not addressed the question of where the conscious experience of inner speech comes from. Levelt (1983) suggested that inner speech represented a perceptual loop where speech production plans were fed into the speech comprehension system. We propose that, although such loops may occur, they may not be necessary for the experience of inner speech. Instead, we associate inner speech directly with phoneme selection (cf. MacKay, 1990). Since phoneme selection is a resource-demanding process (Cook & Meyer, 2008; cf. Ferreira & Pashler, 2002), one might reasonably posit that either the process or its outcome could give rise to conscious awareness without requiring a perceptual loop.⁷

Implications for Other Models of Speech Production

As we have already noted, perceptual loop theories (e.g., Levelt, 1989; Wheeldon & Levelt, 1995) do not have the flexibility to account for the fact that similarity ef-

fects are stronger in silently articulated speech. Since they assume that speech is monitored only via a fixed inner loop and an overt auditory loop and since they lack the potential for subsequent processing to affect the phonological loop, they can assume only that mouthed speech and unmouthed inner speech should be identical. Future revisions of the perceptual loop theory will need to account for the variation. For example, it could be proposed that the inner perceptual processes could flexibly monitor either the phonological level or the articulatory plan, depending on whether articulation is present (e.g., Oomen & Postma, 2001).

Critically, perceptual loop theories also assume that unarticulated inner speech is identical to the inner speech that precedes overt speech—the hypothesized basis for the prearticulatory monitoring that is central to some accounts of speech error patterns (e.g., Levelt et al., 1999; but see Huettig & Hartsuiker, 2010, and Vigliocco & Hartsuiker, 2002, for previous criticisms of this assumption). Many researchers (e.g., Özdemir, Roelefs, & Levelt, 2007; Schiller, Jansma, Peters, & Levelt, 2006; Wheeldon & Levelt, 1995) have used this assumption to characterize prearticulatory monitoring by investigating unarticulated inner speech. If, as we have argued, articulation changes inner speech, this may affect our interpretation of those experiments. Specifically, we cannot afford to ignore the potential contributions of additional articulatory planning.

CONCLUSION

Inner speech is dynamic. It is specified to at least the phoneme level, providing the phonological character that so many researchers report. But when speakers engage in more detailed articulatory planning, their inner speech reflects that information. Repair biases contribute little to the findings on which we base these conclusions.

What, then, can we make of Watson's (1913) original claim that inner speech depends on articulation? Although inner speech may not require overt movements (even very little ones), we have demonstrated that articulation changes inner speech, and this demonstration implies that inner speech cannot be independent of the movements that a person would use to express it.

AUTHOR NOTE

This research was supported by National Institutes of Health Grants DC 000191 and HD 44458. We thank Robert Hartsuiker, Sieb Nooteboom, Kay Bock, Matt Goldrick, Dan Acheson, Don MacKay, Paul Brocklehurst, Javier Ospina, and an anonymous reviewer for their valuable comments and other contributions. Correspondence concerning this article should be addressed to G. M. Oppenheim, Beckman Institute, University of Illinois, 405 North Mathews Avenue, Urbana, IL 61801 (e-mail: goppenh2@illinois.edu).

REFERENCES

- ABRAMSON, M., & GOLDINGER, S. D. (1997). What the reader's eye tells the mind's ear: Silent reading activates inner speech. *Perception & Psychophysics*, *59*, 1059-1068.
- AZIZ-ZADEH, L., CATTANEO, L., ROCHAT, M., & RIZZOLATTI, G. (2005). Covert speech arrest induced by rTMS over both motor and nonmotor left hemisphere frontal sites. *Journal of Cognitive Neuroscience*, *17*, 928-938. doi:10.1162/0898929054021157
- BAARS, B. J., MOTLEY, M. T., & MACKEY, D. G. (1975). Output editing for lexical status in artificially elicited slips of the tongue. *Journal of Verbal Learning & Verbal Behavior*, *14*, 382-391. doi:10.1016/S0022-5371(75)80017-X
- BADDELEY, A., & WILSON, B. (1985). Phonological coding and short-term memory in patients without speech. *Journal of Memory & Language*, *24*, 490-502. doi:10.1016/0749-596X(85)90041-5
- BAILEY, T. M., & HAHN, U. (2005). Phoneme similarity and confusability. *Journal of Memory & Language*, *52*, 339-362. doi:10.1016/j.jml.2004.12.003
- BARCH, D. M., SABB, F. W., CARTER, C. S., BRAVER, T. S., NOLL, D. C., & COHEN, J. D. (1999). Overt verbal responding during fMRI scanning: Empirical investigations of problems and potential solutions. *NeuroImage*, *10*, 642-657. doi:10.1006/nimg.1999.0500
- BARSALOU, L. W. (1999). Perceptual symbol systems. *Behavioral & Brain Sciences*, *22*, 577-609. doi:10.1017/S0140525X99002149
- BROCKLEHURST, P., & CORLEY, M. (2009, September). *Lexical bias and phonemic similarity: The case of inner speech*. Poster presented at the 15th Annual Conference on Architectures and Mechanisms for Language Processing, Barcelona.
- CAPLAN, D., ROCHON, E., & WATERS, G. S. (1992). Articulatory and phonological determinants of word length effects in span tasks. *Quarterly Journal of Experimental Psychology*, *45A*, 177-192. doi:10.1080/14640749208401323
- COOK, A. E., & MEYER, A. S. (2008). Capacity demands of phoneme selection in word production: New evidence from dual-task experiments. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *34*, 886-899. doi:10.1037/0278-7393.34.4.886
- DELL, G. S. (1978). Slips of the mind. In M. Paradis (Ed.), *The fourth LACUS forum* (pp. 69-75). Columbia, SC: Hornbeam.
- DELL, G. S. (1986). A spreading-activation theory of retrieval in sentence production. *Psychological Review*, *93*, 283-321. doi:10.1037/0033-295X.93.3.283
- DELL, G. S. (1990). Effects of frequency and vocabulary type on phonological speech errors. *Language & Cognitive Processes*, *5*, 313-349. doi:10.1080/01690969008407066
- DELL, G. S., & REPKA, R. J. (1992). Errors in inner speech. In B. J. Baars (Ed.), *Experimental slips and human error: Exploring the architecture of volition* (pp. 237-262). New York: Plenum.
- FERREIRA, V. S., & PASHLER, H. (2002). Central bottleneck influences on the processing stages of word production. *Journal of Experimental Psychology: Learning, Memory, & Cognition*, *28*, 1187-1199. doi:10.1037/0278-7393.28.6.1187
- FRISCH, S. (1996). *Similarity and frequency in phonology*. Unpublished doctoral dissertation, Northwestern University, Evanston, IL.
- GALANTUCCI, B., FOWLER, C. A., & TURVEY, M. T. (2006). The motor theory of speech perception reviewed. *Psychonomic Bulletin & Review*, *13*, 361-377.
- GARNSEY, S. M., & DELL, G. S. (1984). Some neurolinguistic implications of prearticulatory editing in production. *Brain & Language*, *23*, 64-73. doi:10.1016/0093-934X(84)90006-3
- GIBBONS, J. D. (1985). *Nonparametric methods for quantitative analysis* (2nd ed.). Columbus, OH: American Sciences Press.
- GOLDRICK, M. (2004). Phonological features and phonotactic constraints in speech production. *Journal of Memory & Language*, *51*, 586-603. doi:10.1016/j.jml.2004.07.004
- GOLDRICK, M., & RAPP, B. (2007). Lexical and post-lexical phonological representations in spoken production. *Cognition*, *102*, 219-260. doi:10.1016/j.cognition.2005.12.010
- HARTSUIKER, R. J. (2006). Are speech error patterns affected by a monitoring bias? *Language & Cognitive Processes*, *21*, 856-891. doi:10.1080/01690960600824138
- HARTSUIKER, R. J., CORLEY, M., & MARTENSEN, H. E. (2005). The lexical bias effect is modulated by context, but the standard monitoring account doesn't fly: Related reply to Baars et al. (1975). *Journal of Memory & Language*, *52*, 58-70. doi:10.1016/j.jml.2004.07.006
- HARTSUIKER, R. J., KOLK, H. H. J., & MARTENSEN, H. (2005). The division of labor between internal and external speech monitoring. In R. J. Hartsuiker, R. Bastiaanse, A. Postma, & F. Wijnen (Eds.), *Phono-*

- logical encoding and monitoring in normal and pathological speech (pp. 187-205). Hove, U.K.: Psychology Press.
- HOLLANDER, M., & WOLFE, D. A. (1973). *Nonparametric statistical methods*. New York: Wiley.
- HU, X., & BATCHELDER, W. H. (1994). The statistical analysis of general processing tree models with the EM algorithm. *Psychometrika*, **59**, 21-47. doi:10.1007/BF02294263
- HUETTIG, F., & HARTSUIKER, R. J. (2010). Listening to yourself is like listening to others: External, but not internal, verbal self-monitoring is based on speech perception. *Language & Cognitive Processes*, **25**, 347-374. doi:10.1080/01690960903046926
- INDEFREY, P., & LEVELT, W. J. M. (2004). The spatial and temporal signatures of word production components. *Cognition*, **92**, 101-144. doi:10.1016/j.cognition.2002.06.001
- KITTREDGE, A. K., DELL, G. S., VERKUILEN, J., & SCHWARTZ, M. F. (2008). Where is the effect of frequency in word production? Insights from aphasic picture-naming errors. *Cognitive Neuropsychology*, **25**, 463-492. doi:10.1080/02643290701674851
- KUCERA, H., & FRANCIS, W. N. (1967). *Computational analysis of present-day American English*. Providence, RI: Brown University Press.
- LAKOFF, G. (1987). *Women, fire, and dangerous things: What categories reveal about the mind*. Chicago: University of Chicago Press.
- LEVELT, W. J. M. (1983). Monitoring and self-repair in speech. *Cognition*, **14**, 41-104. doi:10.1016/0010-0277(83)90026-4
- LEVELT, W. J. M. (1989). *Speaking: From intention to articulation*. Cambridge, MA: MIT Press.
- LEVELT, W. J. M. (1999). Producing spoken language: A blueprint of the speaker. In C. M. Brown & P. Hagoort (Eds.), *The neurocognition of language* (pp. 83-122). Oxford: Oxford University Press.
- LEVELT, W. J. M., ROELOFS, A., & MEYER, A. S. (1999). A theory of lexical access in speech production. *Behavioral & Brain Sciences*, **22**, 1-38. doi:10.1017/S0140525X99001776
- LIBERMAN, A. M., DELATTRE, P., & COOPER, F. S. (1952). The role of selected stimulus-variables in the perception of the unvoiced stop consonants. *American Journal of Psychology*, **65**, 497-516. doi:10.2307/1418032
- LIBERMAN, A. M., & MATTINGLY, I. G. (1985). The motor theory of speech perception revised. *Cognition*, **21**, 1-36. doi:10.1016/0010-0277(85)90021-6
- MACKAY, D. G. (1970). Spoonerisms: The structure of errors in the serial order of speech. *Neuropsychologia*, **8**, 323-350. doi:10.1016/0028-3932(70)90078-3
- MACKAY, D. G. (1981). The problem of rehearsal or mental practice. *Journal of Motor Behavior*, **13**, 274-285.
- MACKAY, D. G. (1990). Perception, action, and awareness: A three-body problem. In O. Neumann, & W. Prinz (Eds.), *Relationships between perception and action* (pp. 269-303). Heidelberg: Springer.
- MACKAY, D. G. (1992). Constraints on theories of inner speech. In D. Reisberg (Ed.), *Auditory imagery* (pp. 121-149). Hillsdale, NJ: Erlbaum.
- MCGUIRE, P. K., SILBERSWEIG, D. A., MURRAY, R. M., DAVID, A. S., FRACKOWIAK, R. S., & FRITH, C. D. (1996). Functional anatomy of inner speech and auditory verbal imagery. *Psychological Medicine*, **26**, 29-38. doi:10.1017/S0033291700033699
- MOSHAGEN, M. (2010). multiTree: A computer program for the analysis of multinomial processing tree models. *Behavior Research Methods*, **42**, 42-54. doi:10.3758/BRM.42.1.42
- NICKELS, L., & HOWARD, D. (1995). Phonological errors in aphasic naming: Comprehension, monitoring and lexicality. *Cortex*, **31**, 209-237.
- NOOTEBOOM, S. G. (2005). Lexical bias revisited: Detecting, rejecting and repairing speech errors in inner speech. *Speech Communication*, **47**, 43-58. doi:10.1016/j.specom.2005.02.003
- NOOTEBOOM, S., & QUENÉ, H. (2008). Self-monitoring and feedback: A new attempt to find the main cause of lexical bias in phonological speech errors. *Journal of Memory & Language*, **58**, 837-861. doi:10.1016/j.jml.2007.05.003
- NOZARI, N., & DELL, G. S. (2009). More on lexical bias: How efficient can a "lexical editor" be? *Journal of Memory & Language*, **60**, 291-307. doi:10.1016/j.jml.2008.09.006
- OOMEN, C. C., & POSTMA, A. (2001). Effects of time pressure on mechanisms of speech production and self-monitoring. *Journal of Psycholinguistic Research*, **30**, 163-184. doi:10.1023/A:1010377828778
- OPPENHEIM, G. M., & DELL, G. S. (2008). Inner speech slips exhibit lexical bias, but not the phonemic similarity effect. *Cognition*, **106**, 528-537. doi:10.1016/j.cognition.2007.02.006
- ÖZDEMİR, R., ROELOFS, A., & LEVELT, W. J. (2007). Perceptual uniqueness point effects in monitoring internal speech. *Cognition*, **105**, 457-465. doi:10.1016/j.cognition.2006.10.006
- PALMER, E. D., ROSEN, H. J., OJEMANN, J. G., BUCKNER, R. L., KELLEY, W. M., & PETERSEN, S. E. (2001). An event-related fMRI study of overt and covert word stem completion. *NeuroImage*, **14**, 182-193. doi:10.1006/nimg.2001.0779
- PAULESU, E., FRITH, C. D., & FRACKOWIAK, R. S. J. (1993). The neural correlates of the verbal component of working memory. *Nature*, **362**, 342-345. doi:10.1038/362342a0
- POSTMA, A. (2000). Detection of errors during speech production: A review of speech monitoring models. *Cognition*, **77**, 97-132. doi:10.1016/S0010-0277(00)00090-1
- POSTMA, A., & NOORDANUS, C. (1996). The production and detection of speech errors in silent, mouthed, noise-masked, and normal auditory feedback speech. *Language & Speech*, **39**, 375-392.
- PULVERMÜLLER, F. (2005). Brain mechanisms linking language and action. *Nature Reviews Neuroscience*, **6**, 576-582. doi:10.1038/nrn1706
- REISBERG, D., SMITH, J. D., BAXTER, D. A., & SONENSHINE, M. (1989). "Enacted" auditory images are ambiguous; "pure" auditory images are not. *Quarterly Journal of Experimental Psychology*, **41A**, 619-641.
- ROELOFS, A. (2004). Error biases in spoken word planning and monitoring by aphasic and nonaphasic speakers: Comment on Rapp and Goldrick (2000). *Psychological Review*, **111**, 561-572. doi:10.1037/0033-295X.111.2.561
- SAHIN, N. T., PINKER, S., CASH, S. S., SCHOMER, D., & HALGREN, E. (2009). Sequential processing of lexical, grammatical, and phonological information within Broca's area. *Science*, **326**, 445-449. doi:10.1126/science.1174481
- SCHILLER, N. O., JANSMA, B. M., PETERS, J., & LEVELT, W. J. M. (2006). Monitoring metrical stress in polysyllabic words. *Language & Cognitive Processes*, **21**, 112-140. doi:10.1080/01690960400001861
- SHERGILL, S. S., BULLMORE, E. T., BRAMMER, M. J., WILLIAMS, S. C. R., MURRAY, R. M., & MCGUIRE, P. K. (2001). A functional study of auditory verbal imagery. *Psychological Medicine*, **31**, 241-253. doi:10.1017/S003329170100335X
- SHEKIN, D. (2000). *Handbook of parametric and nonparametric statistical procedures* (2nd ed.). Boca Raton, FL: Chapman & Hall/CRC Press.
- SHUSTER, L. I., & LEMIEUX, S. K. (2005). An fMRI investigation of covertly and overtly produced mono- and multisyllabic words. *Brain & Language*, **93**, 20-31. doi:10.1016/j.bandl.2004.07.007
- SMITH, J. D., REISBERG, D., & WILSON, M. (1992). Subvocalization and auditory imagery: Interactions between the inner ear and the inner voice. In D. Reisberg (Ed.), *Auditory imagery* (pp. 95-119). Hillsdale, NJ: Erlbaum.
- SMITH, S. M., BROWN, H. O., TOMAN, J. E. P., & GOODMAN, L. S. (1947). The lack of cerebral effects of d-tubocuarine. *Anesthesiology*, **8**, 1-14. doi:10.1097/0000542-194701000-00001
- SOKOLOV, A. N. (1972). *Inner speech and thought*. New York: Plenum.
- STEMBERGER, J. P. (1991). Apparent anti-frequency effects in language production: The addition bias and phonological underspecification. *Journal of Memory & Language*, **30**, 161-185. doi:10.1016/0749-596X(91)90002-2
- THOMPSON-SCHILL, S. L., D'ESPOSITO, M., AGUIRRE, G. K., & FARAH, M. J. (1997). Role of left inferior prefrontal cortex in retrieval of semantic knowledge: A reevaluation. *Proceedings of the National Academy of Sciences*, **94**, 14792-14797. doi:10.1073/pnas.94.26.14792
- VALLAR, G., & CAPP, S. F. (1987). Articulation and verbal short-term memory: Evidence from anarthria. *Cognitive Neuropsychology*, **4**, 55-77. doi:10.1080/02643298708252035
- VIGLIOCO, G., & HARTSUIKER, R. J. (2002). The interplay of meaning, sound, and syntax in sentence production. *Psychological Bulletin*, **128**, 442-472. doi:10.1037/0033-2909.128.3.442
- VYGOTSKIĬ, L. S. (1965). *Thought and language*. Cambridge, MA: MIT Press.

- WATSON, J. B. (1913). Psychology as the behaviorist views it. *Psychological Review*, **20**, 158-177. doi:10.1037/h0074428
- WHEELDON, L. R., & LEVELT, W. J. M. (1995). Monitoring the time course of phonological encoding. *Journal of Memory & Language*, **34**, 311-334. doi:10.1006/jmla.1995.1014
- YETKIN, F. Z., HAMMEKE, T. A., SWANSON, S. J., MORRIS, G. L., MUELLER, W. M., MCAULIFFE, T. L., & HAUGHTON, V. M. (1995). A comparison of functional MR activation patterns during silent and audible language tasks. *American Journal of Neuroradiology*, **16**, 1087-1092.

NOTES

1. Levelt et al. (1999) appeared rather noncommittal about the units of inner speech—whether it consists of a phonological code, a phonetic code, or both—but the “Authors’ Response” (p. 64) and their concurrent and subsequent descriptions of the theory (e.g., Indefrey & Levelt, 2004; Levelt, 1999; Özdemir, Roelofs, & Levelt, 2007; Roelofs, 2004) indicate a shift from phonetic to phonological conceptions of inner speech.
2. We replaced one item set that showed a strong reverse phonemic similarity effect in Oppenheim and Dell’s (2008) overt condition, suggesting a possible limitation of our three-feature metric for assessing similarity (cf. Frisch, 1996). However, we did include the item as a filler trial and found that including it in our analyses would not have changed the outcomes of our statistical analyses.

3. This place–manner–voicing similarity metric offers the benefit of being easy to use while still providing good predictive power (e.g., Bailey & Hahn, 2005). Moreover, our resulting phoneme contrasts agree with those arrived at by more complex measures (i.e., Frisch’s, 1996, natural classes metric) for 31 out of 32 items. Excluding the aberrant contrast did not change the outcome of any statistical test reported herein.

4. Two nonword outcomes (*beel* and *hoak*) corresponded to very low frequency entries in Kučera and Francis’s (1967) analysis.

5. Note that applying this method to self-reported errors will simultaneously test the hypothesis that participants may misperceive or misreport their errors.

6. Because of identifiability constraints, this proportion is necessarily fixed. The value is admittedly arbitrary, but we note that changing this parameter does not alter any of the effects that we describe.

7. Note that assigning awareness of inner speech more directly to the production system could explain a puzzling set of findings: Although speakers can hear their inner speech and appear to monitor their inner speech when speaking aloud (e.g., Levelt, 1983), their eye movements suggest that they do not actually *listen* to their inner speech while speaking aloud (Huettig & Hartsuiker, 2010).

(Manuscript received August 23, 2009;
revision accepted for publication May 7, 2010.)